Parameter Estimation with Neural Simulation-Based Inference in ATLAS

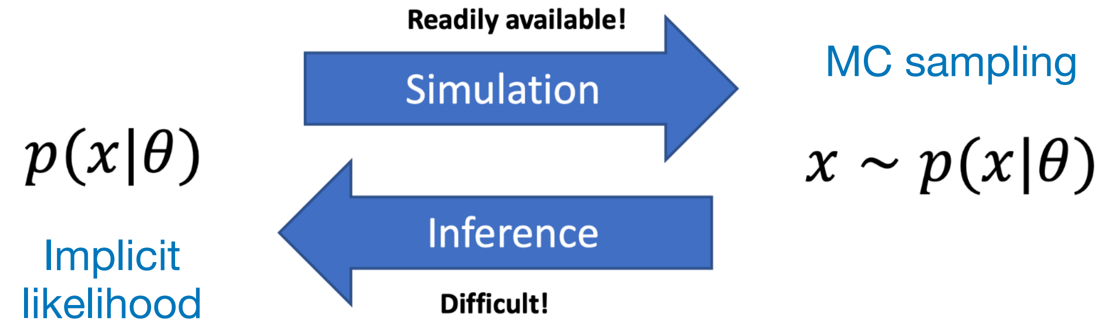
Jay Sandesara
on behalf of the ATLAS collaboration

UNIVERSITY OF WISCONSIN–MADISON

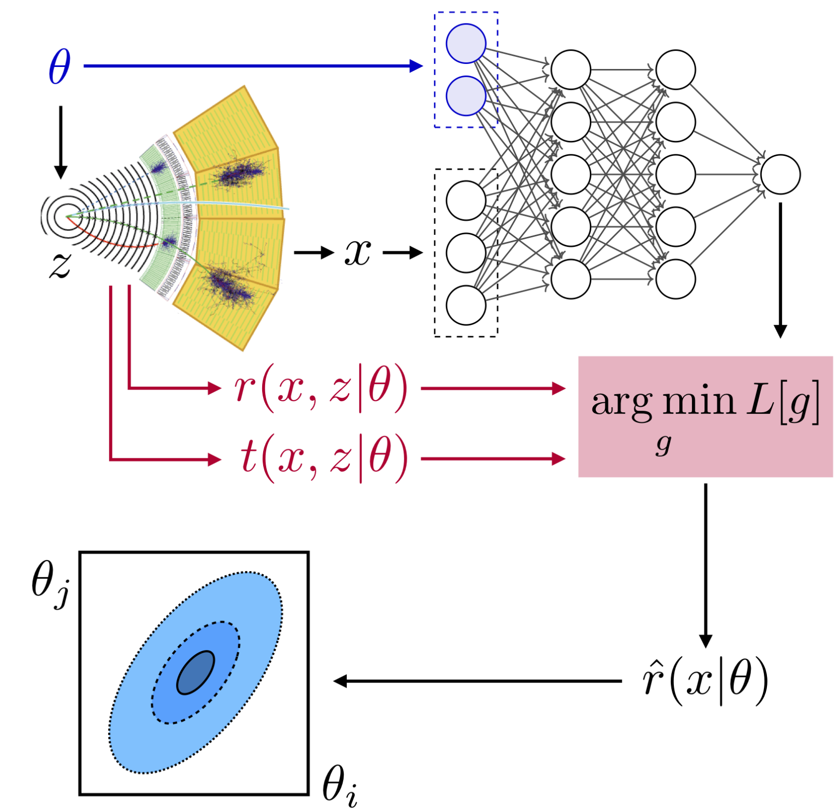
ATLAS EXPERIMENT

Introduction

- Neural Simulation-Based Inference (NSBI) covers a broad range of statistical techniques.



- **Idea:** build ML surrogates for powerful statistical inference in the presence of
    - Intractable likelihoods (e.g. LHC analysis), or
    - when likelihoods are slow to compute analytically (e.g. gravitational wave analysis).



Overview of typical NSBI workflow

Figure credits: Madminer [1805.00020, 1805.00013, 1805.12244]

Comprehensive review of NSBI: 1911.01429

NSBI at the LHC

- The focus of this talk is on a **practical application of these methods to LHC analysis**. The talk will cover:

  - Efficiently modelling likelihoods as a function of **complex high-dimensional parameter space.** ✓

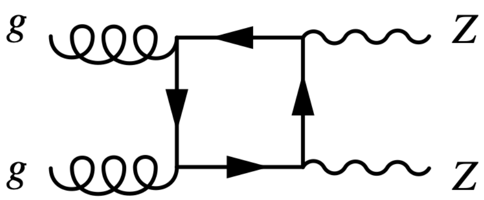
  - Rigorously **testing the quality of the surrogate models** and their reliability using MC and real data. ✓

  - Building **robust frequentist confidence intervals** using Neyman Construction. ✓

- This will be explained with an example of the off-shell Higgs boson measurement at the ATLAS experiment [Rep. Prog. Phys. 88 067801, Rep. Prog. Phys. 88 057803]

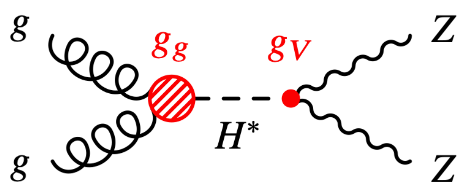
NSBI at the LHC

Parton-level events sampled from analytical model



$z \sim p_B(z \,|\, \mu)$

$z \sim p_S(z \,|\, \mu)$



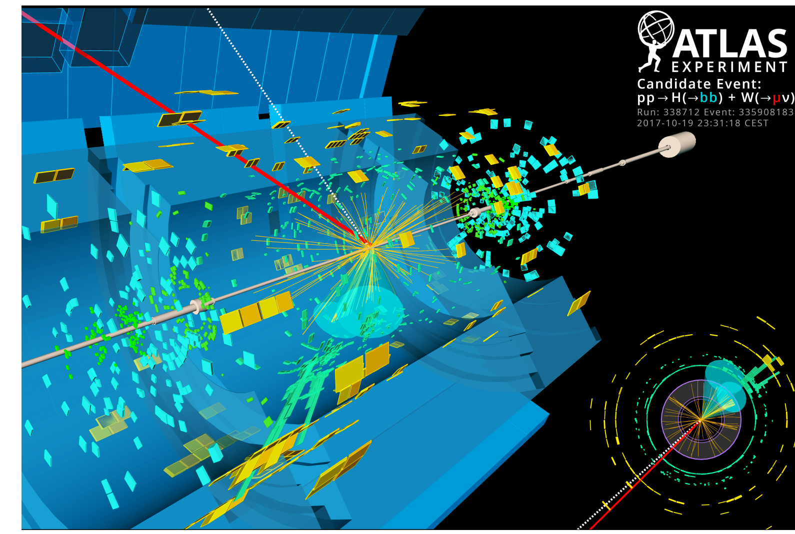
Parton Shower

Hadronization

Detector Reconstruction

MC events sampled from implicit likelihoods

$$x \sim p_S(x \,|\, \mu) \,, p_B(x \,|\, \mu)$$



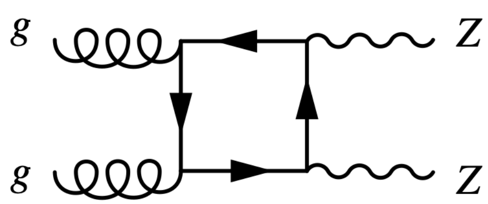
Simulation (forward pass)

NSBI at the LHC

MC events sampled from implicit likelihoods

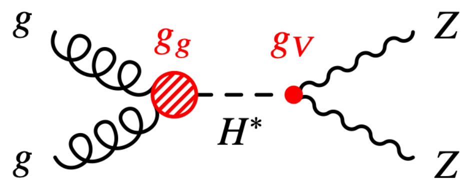
$$x \sim p_S(x \mid \mu)\,,\, p_B(x \mid \mu)$$

Parton-level events sampled from analytical model



$z \sim p_B(z \mid \mu)$

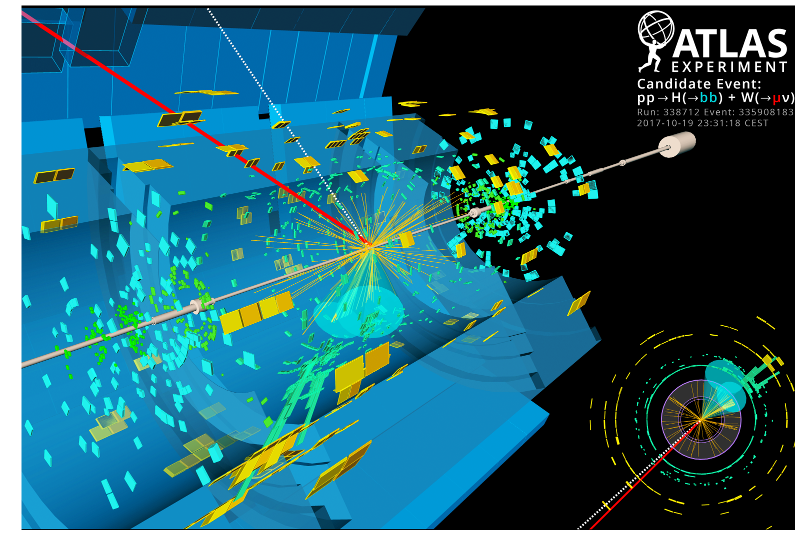
$z \sim p_S(z \mid \mu)$



Parton Shower

Hadronization

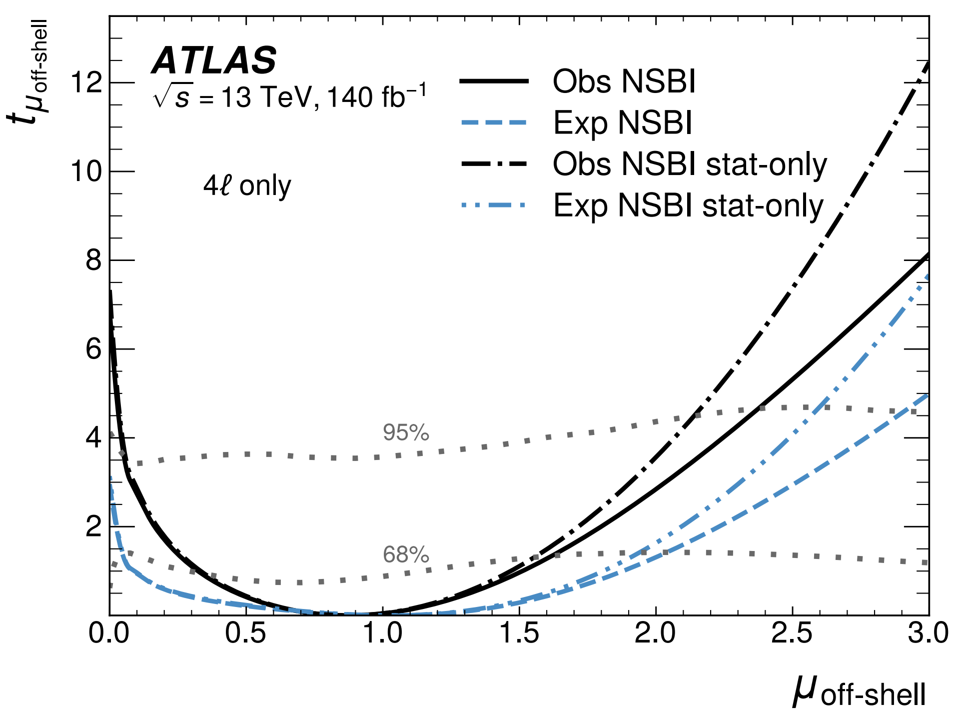
Detector Reconstruction



Inference (reverse pass)

Reconstructed events x_i

Inference on model parameter μ



Profile Negative Log-Likelihood Fit

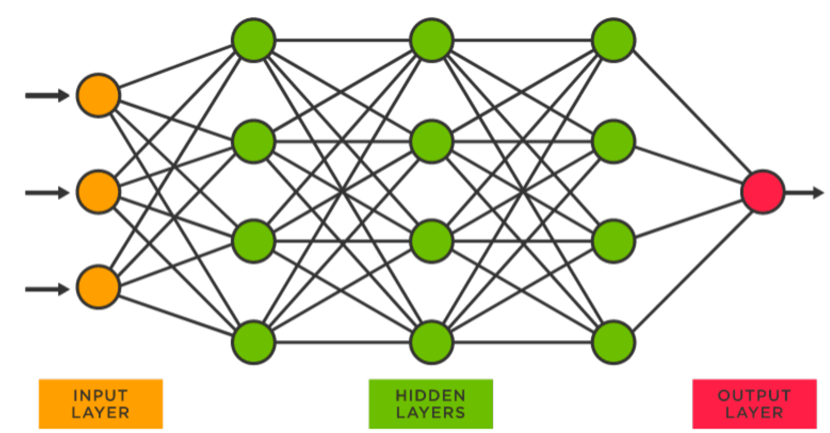
Rep. Prog. Phys. 88 057803

$$-2 \cdot \sum_{i \in events} \log \frac{p(x_i \mid \mu, \hat{\hat{\alpha}})}{p(x_i \mid \hat{\mu}, \hat{\alpha})}$$

Event-by-event parameterized likelihood ratios

$\mu \rightarrow$ parameter of interest

$\alpha \rightarrow$ nuisance parameters



Surrogate Model for likelihood ratios

The off-shell Higgs boson

The probability model of the off-shell Higgs boson:

$$p(x\,|\,\mu) = \frac{1}{\nu(\mu)}\left[\underbrace{\mu}\cdot\nu_S\cdot p_S(x) + \underbrace{\sqrt{\mu}}\cdot\nu_I\cdot p_I(x) + \nu_B\cdot p_B(x) + \nu_{NI}\cdot p_{NI}(x)\right]$$

$\nu \rightarrow$ Exp events

Parameter dependence

$$\mu = \frac{\sigma_{obs}^{H\rightarrow ZZ}}{\sigma_{exp}^{H\rightarrow ZZ}}$$

NI $\rightarrow$ Non-Interfering backgrounds

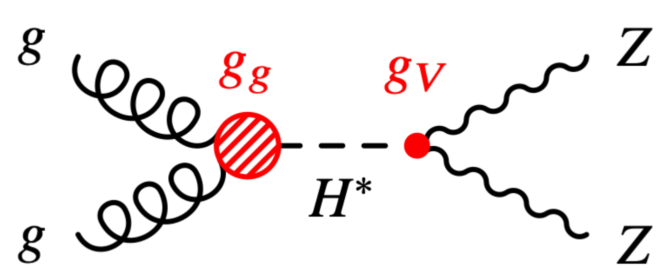
Signal probability

$p_S(x)$

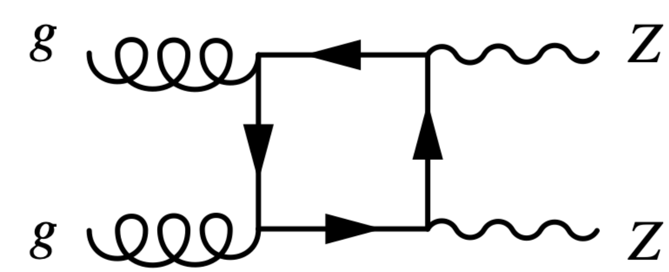
Bkg probability

$p_B(x)$

$p_I(x) = \; 2\cdot Re \; \Big|$



$\times$



$\Big|$

Interference probability

ggF Higgs Signal

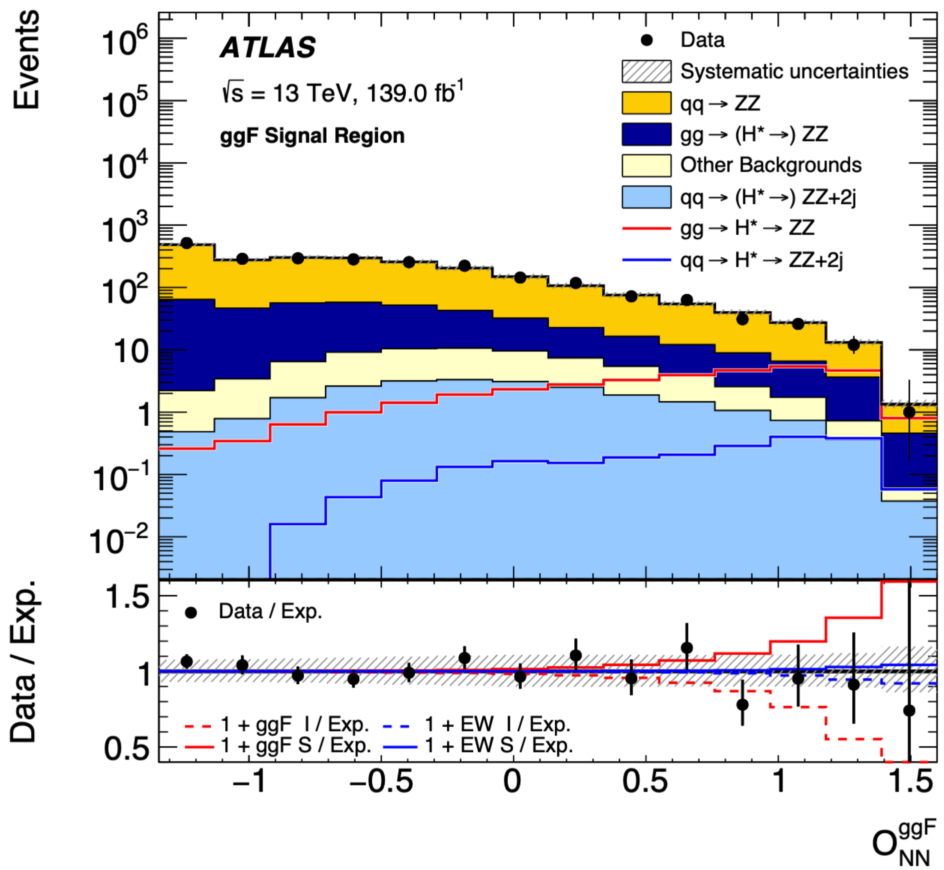
ggF Interfering Background

Previous Analysis

Fixed S/B discriminant is often the optimal choice for hypothesis testing at the LHC

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_B(x)}} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

Previous off-shell analysis
Phys. Lett. B 846 (2023) 138223



$$O_{NN}(x) = \log \frac{p_S(x)}{p_B(x) + 0.1 \cdot p_{NI}(x)}$$

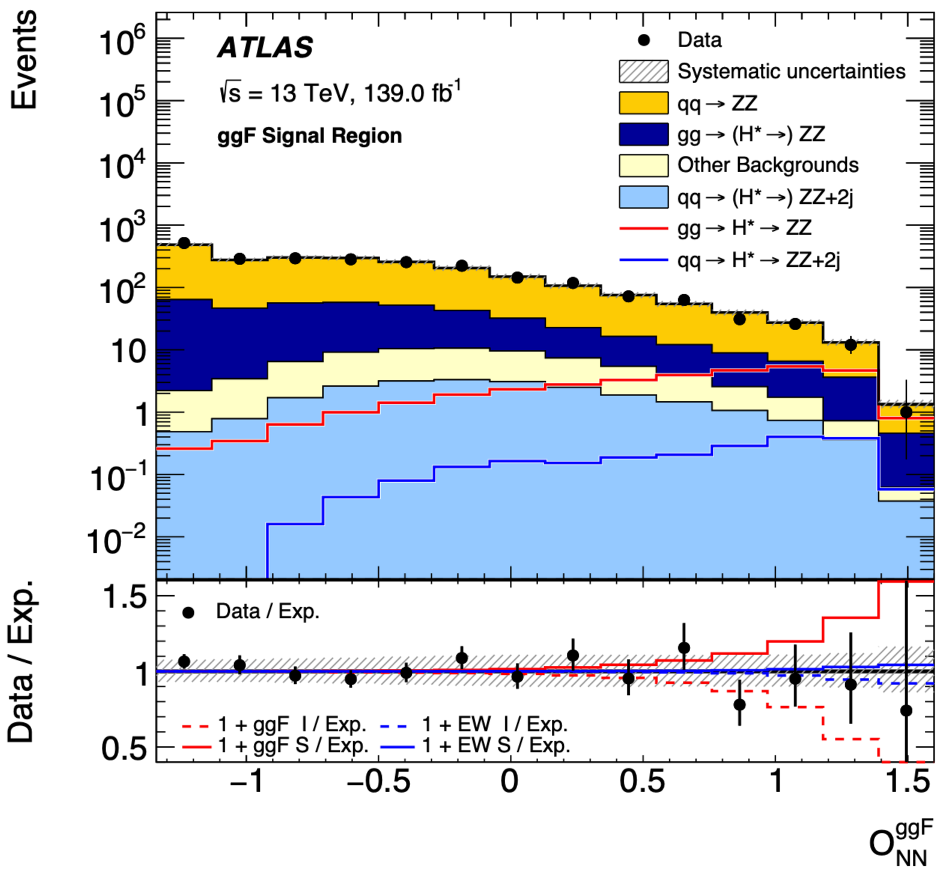
S/B discriminant used in previous Run-2 analysis of off-shell Higgs boson

Previous Analysis

Fixed S/B discriminant is often the optimal choice for hypothesis testing at the LHC

$$\frac{p(x\,|\,\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_B(x)}} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

Previous off-shell analysis
Phys. Lett. B 846 (2023) 138223



Trained using multi-dimensional input feature space

$$O_{NN}(x) = \log \frac{p_S(x)}{p_B(x) + 0.1 \cdot p_{NI}(x)}$$

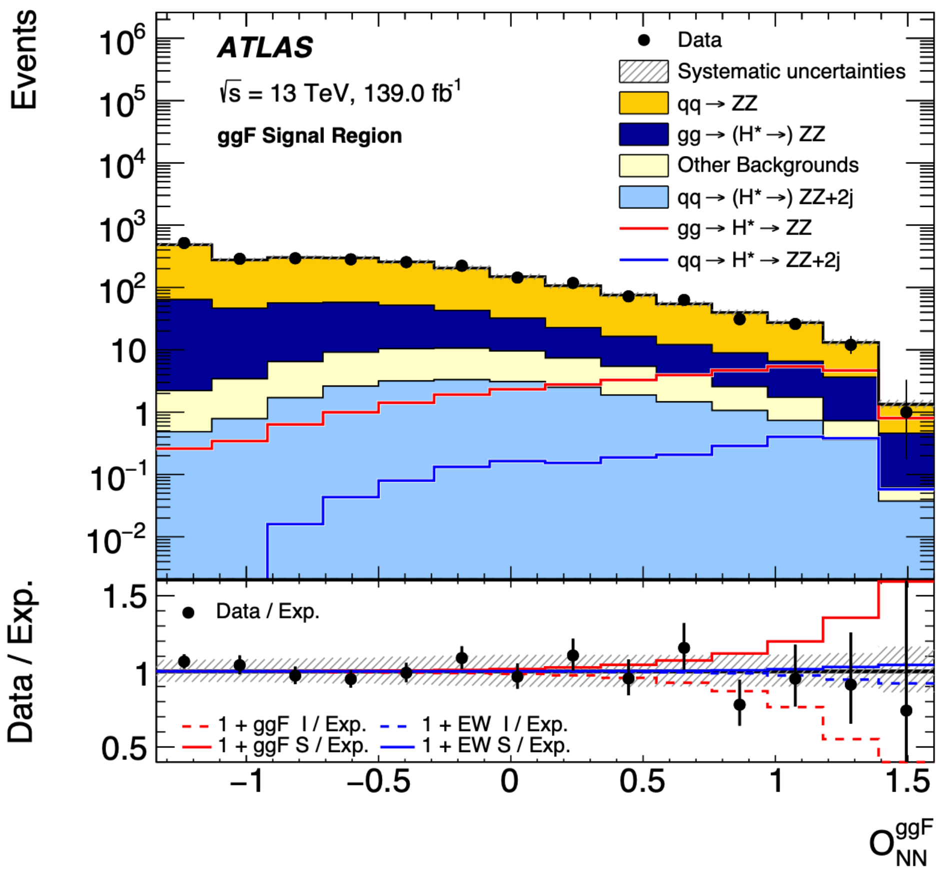
S/B discriminant used in previous Run-2 analysis of off-shell Higgs boson

Previous Analysis

Fixed S/B discriminant is often the optimal choice for hypothesis testing at the LHC

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

Previous off-shell analysis
Phys. Lett. B 846 (2023) 138223



$$O_{NN}(x) = \log \frac{p_S(x)}{p_B(x) + 0.1 \cdot p_{NI}(x)}$$

But what if the parameterization is non-linear?

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \sqrt{\mu} \cdot \nu_I \cdot \underbrace{\frac{p_I(x)}{p_B(x)}} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

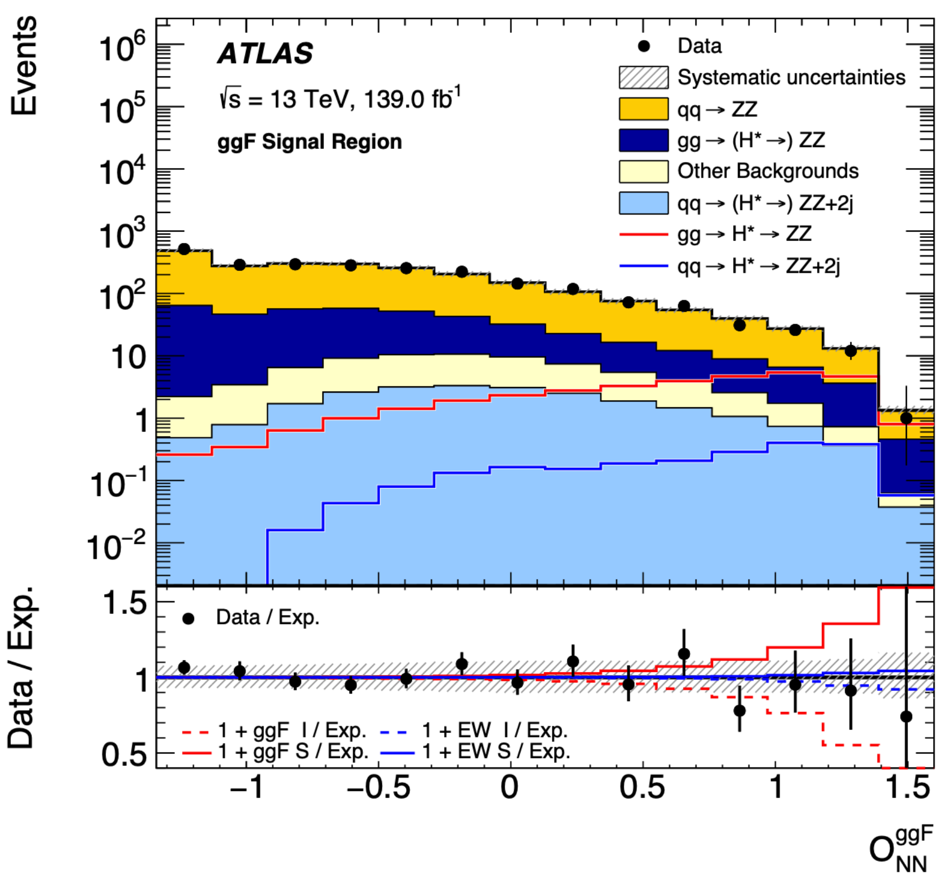
E.g.: interference effects of off-shell Higgs boson production. Single observable no longer describes the full parameter space!

Previous Analysis

But what if the parameterization is non-linear?

Fixed S/B discriminant is often the optimal choice for hypothesis testing at the LHC

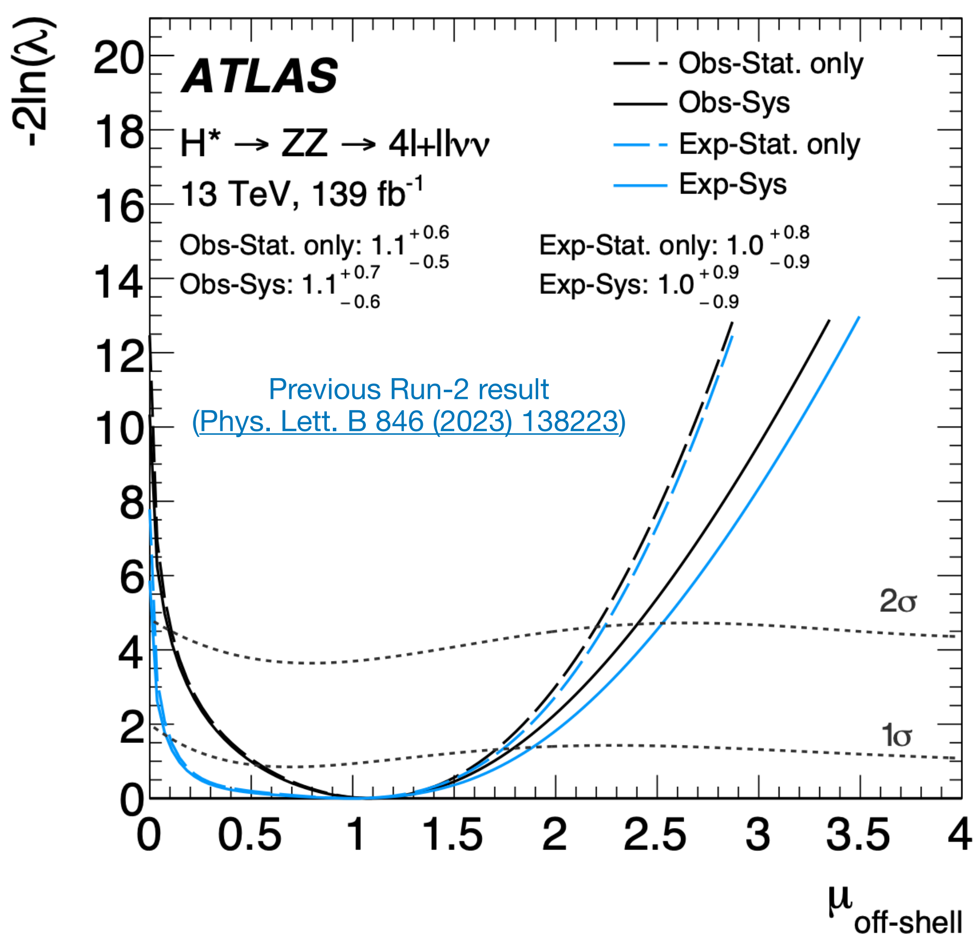
$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$



$$O_{NN}(x) = \log \frac{p_S(x)}{p_B(x) + 0.1 \cdot p_{NI}(x)}$$

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \sqrt{\mu} \cdot \nu_I \cdot \frac{p_I(x)}{p_B(x)} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

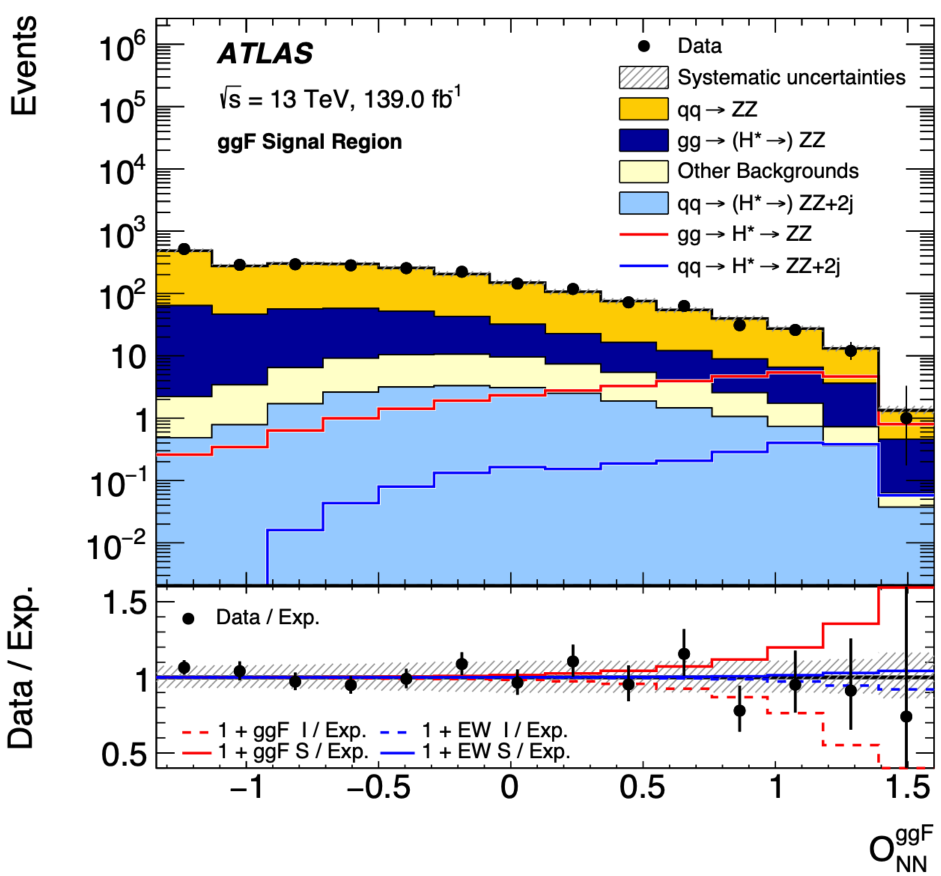
E.g.: interference effects of off-shell Higgs boson production. Single observable no longer describes the full parameter space!



Previous Analysis

Fixed S/B discriminant is often the optimal choice for hypothesis testing at the LHC

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

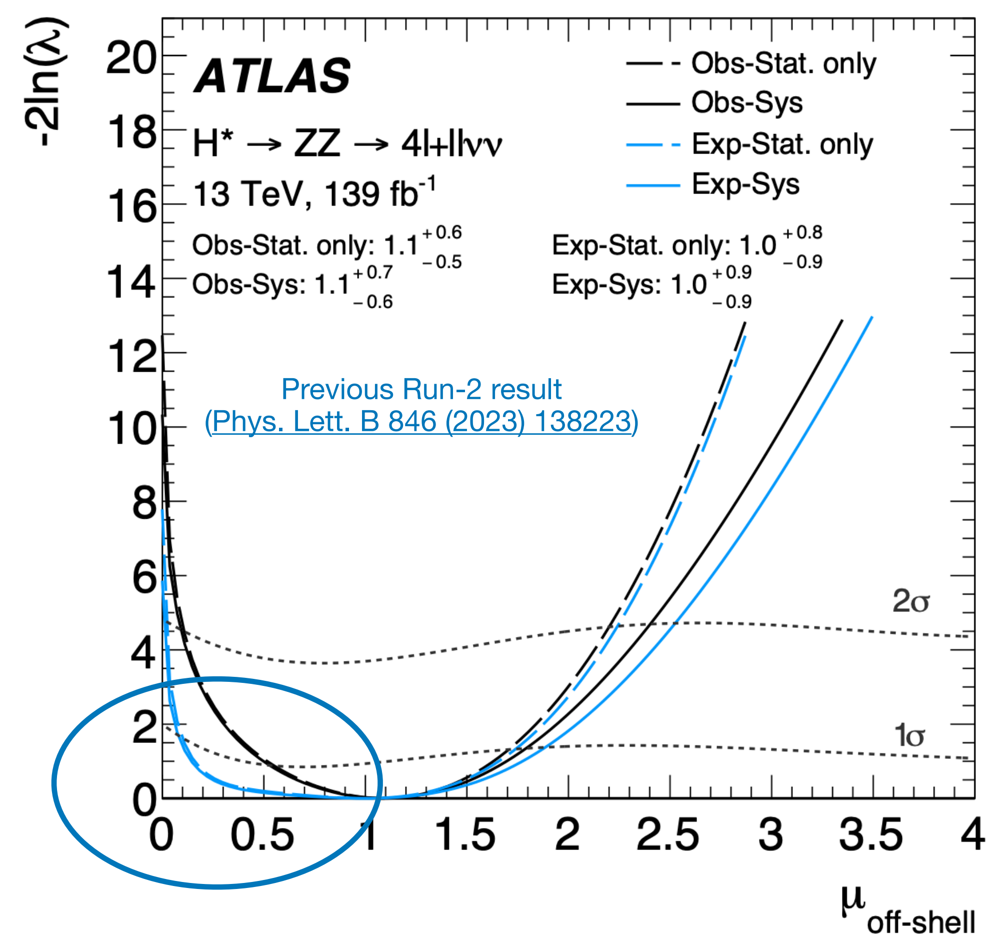


$$O_{NN}(x) = \log \frac{p_S(x)}{p_B(x) + 0.1 \cdot p_{NI}(x)}$$

But what if the parameterization is non-linear?

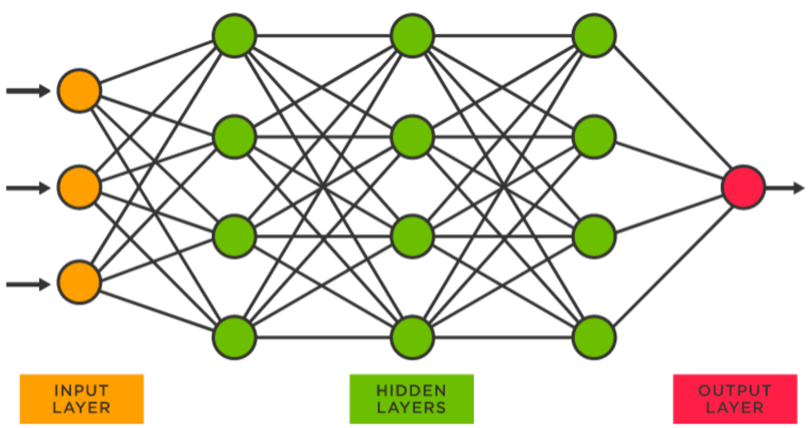
$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \frac{p_S(x)}{p_B(x)} + \sqrt{\mu} \cdot \nu_I \cdot \frac{p_I(x)}{p_B(x)} + \nu_B \cdot \frac{p_B(x)}{p_B(x)}$$

E.g.: interference effects of off-shell Higgs boson production. Single observable no longer describes the full parameter space!



Flat NLL region implies sub-optimality in regions with $\sqrt{\mu} \cdot \nu_I \cdot p_I \gg \mu \cdot \nu_S \cdot p_S$

New Analysis - NSBI



Surrogate Model for likelihood ratios

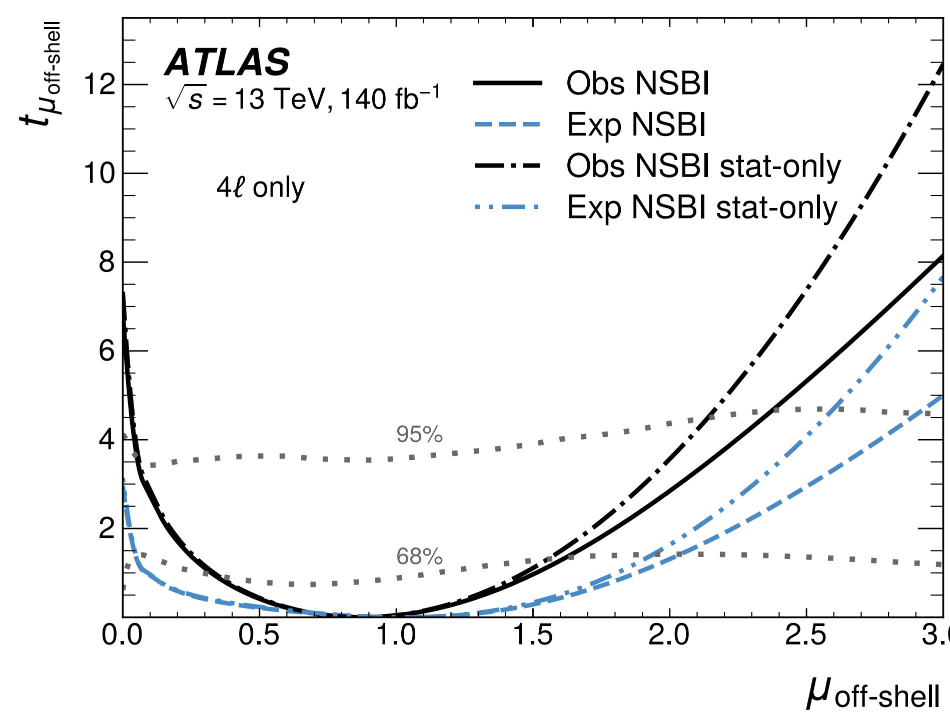
MC events sampled from implicit likelihoods

$$x \sim p_S(x\,|\,\mu)\,,p_B(x\,|\,\mu)$$

Frequentist test using NSBI

Profile Negative Log-Likelihood Test Statistic

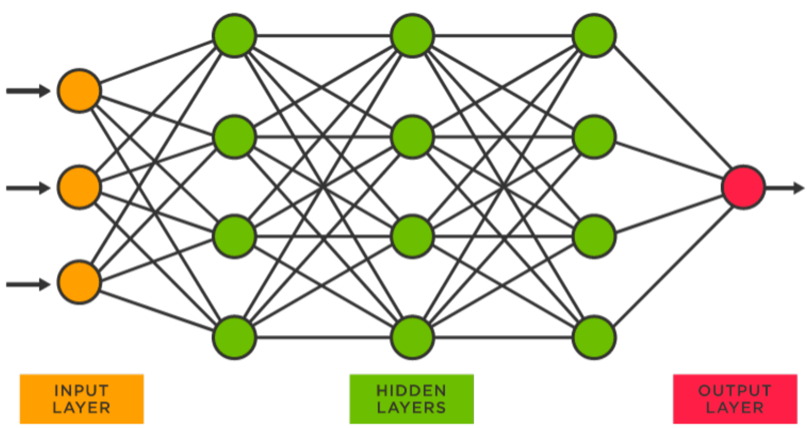
$$-2 \cdot \sum_{i \in events} \log \frac{p(x_i\,|\,\mu,\hat{\hat{\alpha}})}{p(x_i\,|\,\hat{\mu},\hat{\alpha})}$$



Off-shell Higgs measurement using NSBI
Rep. Prog. Phys. 88 057803

New Analysis - NSBI

How do we train this model?



Surrogate Model for likelihood ratios

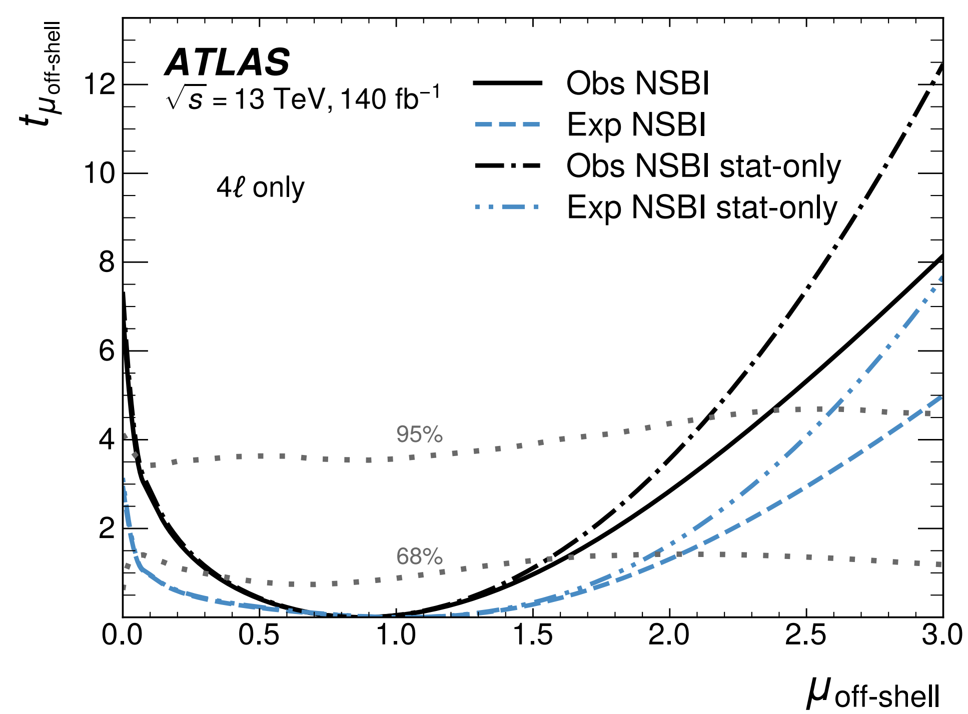
MC events sampled from implicit likelihoods

$$x \sim p_S(x\,|\,\mu)\,,p_B(x\,|\,\mu)$$

Frequentist test using NSBI

Profile Negative Log-Likelihood Test Statistic

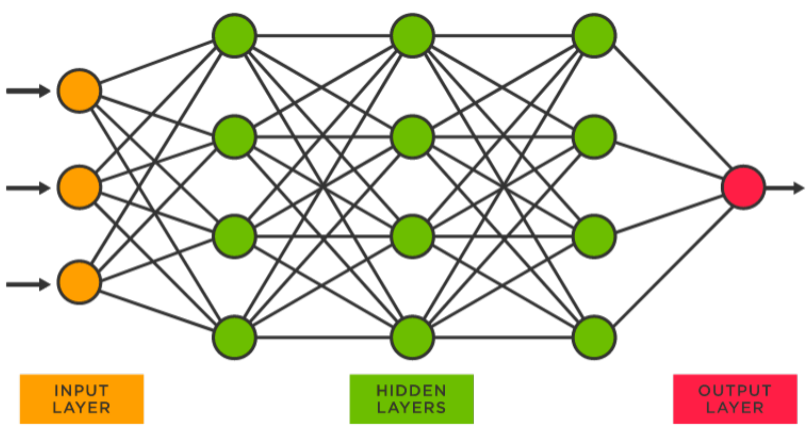
$$-2 \cdot \sum_{i \in events} \log \frac{p(x_i\,|\,\mu, \hat{\hat{\alpha}})}{p(x_i\,|\,\hat{\mu}, \hat{\alpha})}$$



Off-shell Higgs measurement using NSBI
Rep. Prog. Phys. 88 057803

New Analysis - NSBI

How do we train this model?



Surrogate Model for likelihood ratios

MC events sampled from implicit likelihoods

$$x \sim p_S(x\,|\,\mu)\,,p_B(x\,|\,\mu)$$

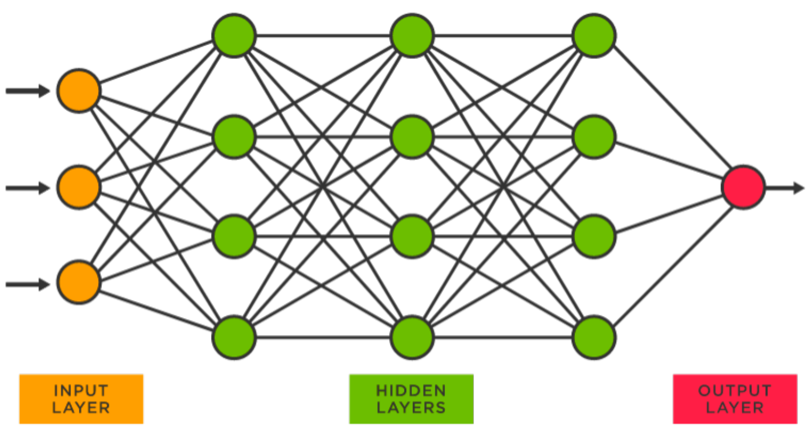
Proposal 1: estimate paramaterized PDFs

$$p(x\,|\,\mu, \alpha)$$

train generative models with tractable probability densities (e.g. Normalizing Flows)

New Analysis - NSBI

How do we train this model?



Surrogate Model for likelihood ratios

MC events sampled from implicit likelihoods

$$x \sim p_S(x \,|\, \mu)\,, p_B(x \,|\, \mu)$$

Proposal 1: estimate paramaterized PDFs

$$p(x \,|\, \mu, \alpha)$$

train generative models with tractable probability densities (e.g. Normalizing Flows)

Proposal 2: Estimate parameterized density ratios

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)}$$

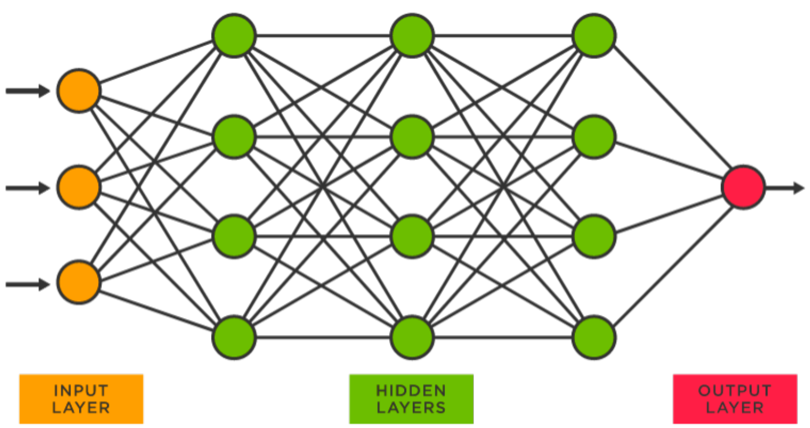
by training well-calibrated and unbiased NN classifiers and use in the profile likelihood ratio:

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) \,/\, p_{ref}(x)}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) \,/\, p_{ref}(x)} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}})}{p(x_i \,|\, \hat{\mu}, \hat{\alpha})}$$

$p_{ref}(x)$ can be any chosen **parameter-independent** hypothesis

New Analysis - NSBI

How do we train this model?



Surrogate Model for likelihood ratios

MC events sampled from implicit likelihoods

$$x \sim p_S(x \,|\, \mu)\,,\, p_B(x \,|\, \mu)$$

Proposal 1: estimate paramaterized PDFs

$$p(x \,|\, \mu, \alpha)$$

train generative models with tractable probability densities (e.g. Normalizing Flows)

Proposal 2: Estimate parameterized density ratios

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)}$$

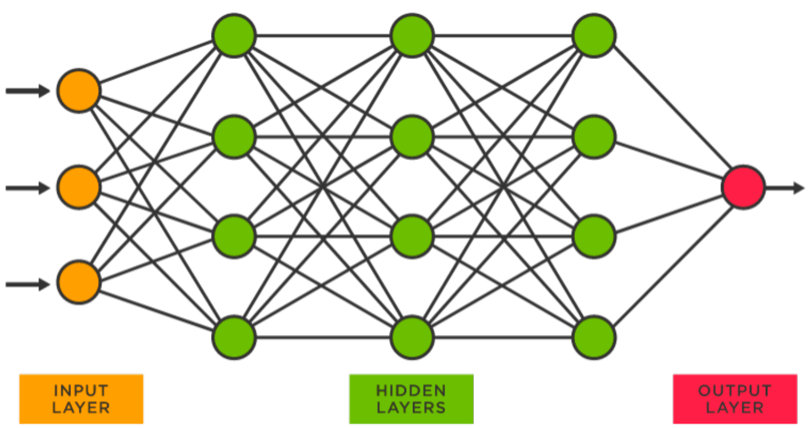
by training well-calibrated and unbiased NN classifiers and use in the profile likelihood ratio:

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) \,/\, \cancel{p_{ref}(x)}}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) \,/\, \cancel{p_{ref}(x)}} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}})}{p(x_i \,|\, \hat{\mu}, \hat{\alpha})}$$

$p_{ref}(x)$ can be any chosen **parameter-independent** hypothesis

New Analysis - NSBI

How do we train this model?



Surrogate Model for likelihood ratios

↑

MC events sampled from implicit likelihoods

$$x \sim p_S(x \,|\, \mu)\,, p_B(x \,|\, \mu)$$

Proposal 1: estimate paramaterized PDFs

$$p(x \,|\, \mu, \alpha)$$

train generative models with tractable probability densities (e.g. Normalizing Flows)

Easier to train and validate for large-dimensional inputs ✔

Proposal 2: Estimate parameterized density ratios

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)}$$

by training well-calibrated and unbiased NN classifiers and use in the profile likelihood ratio:

$$\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x)} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) \,/\, \cancel{p_{ref}(x)}}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) \,/\, \cancel{p_{ref}(x)}} \rightarrow \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}})}{p(x_i \,|\, \hat{\mu}, \hat{\alpha})}$$

$p_{ref}(x)$ can be any chosen **parameter-independent** hypothesis

Overview: Neural Simulation-Based Inference

Full test statistic function for frequentist parameter estimation on parameter μ

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \,|\, \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \,|\, \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}})/\cancel{p_{ref}(x_i)}}{p(x_i \,|\, \hat{\mu}, \hat{\alpha})/\cancel{p_{ref}(x_i)}} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

Extended Poisson term

Sum of event-by-event log-likelihood ratios

Constraint terms

$N_{obs} \rightarrow$ total observed events

$p_{subs} \rightarrow$ likelihood from subsidiary measurements of the nuisance parameters

Overview: Neural Simulation-Based Inference

Full test statistic function for frequentist parameter estimation on parameter μ

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \mid \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \mid \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \mid \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \mid \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

sum over processes
$c = S,\, B,$ **etc.**

parameter-independent ratio

$$\boxed{\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)}} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i \mid \alpha) \cdot \nu_c \cdot \boxed{\frac{p_c(x_i)}{p_{ref}(x_i)}} \right]$$

Parameterized per-event ratios

Overview: Neural Simulation-Based Inference

Full test statistic function for frequentist parameter estimation on parameter μ

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs}\,|\,\mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs}\,|\,\hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i\,|\,\mu, \hat{\hat{\alpha}})/p_{ref}(x_i)}{p(x_i\,|\,\hat{\mu}, \hat{\alpha})/p_{ref}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

sum over processes $c = S, B,$ etc.

parameter-independent ratio

$$\frac{p(x_i\,|\,\mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i\,|\,\alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Parameterized per-event ratios

Parameter dependancies are factorized out

"Mixture models"

E.g. in stat-only off-shell Higgs model:

$$p(x\,|\,\mu) = \frac{1}{\nu(\mu)} \left[\underbrace{\mu}_{f_S(\mu)} \cdot \nu_S \cdot p_S(x) + \underbrace{\sqrt{\mu}}_{f_I(\mu)} \cdot \nu_I \cdot p_I(x) + \nu_B \cdot p_B(x) + \nu_{NI} \cdot p_{NI}(x) \right]$$

Overview: Neural Simulation-Based Inference

Full test statistic function for frequentist parameter estimation on parameter μ

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \mid \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \mid \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \mid \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}(x_i)}}{p(x_i \mid \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}(x_i)}} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

sum over processes $c = S, B,$ **etc.**

parameter-independent ratio

$$\frac{p(x_i \mid \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i \mid \alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Parameterized per-event ratios

Factorized **nuisance parameter** α-dependence:

$$g_c(x \mid \alpha) = \frac{p_c(x \mid \alpha)}{p_c(x)}$$

Parameter dependancies are factorized out

"Mixture models"

Overview: Neural Simulation-Based Inference

Full test statistic function for frequentist parameter estimation on parameter μ

$$t(\mu) = -2 \cdot \log \frac{\text{Pois}(N_{obs} \,|\, \mu, \hat{\hat{\alpha}})}{\text{Pois}(N_{obs} \,|\, \hat{\mu}, \hat{\alpha})} - 2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i \,|\, \mu, \hat{\hat{\alpha}}) / \cancel{p_{ref}}(x_i)}{p(x_i \,|\, \hat{\mu}, \hat{\alpha}) / \cancel{p_{ref}}(x_i)} - 2 \cdot \sum_{k}^{N_{syst}} \log \frac{p_{subs}(\hat{\hat{\alpha}})}{p_{subs}(\hat{\alpha})}$$

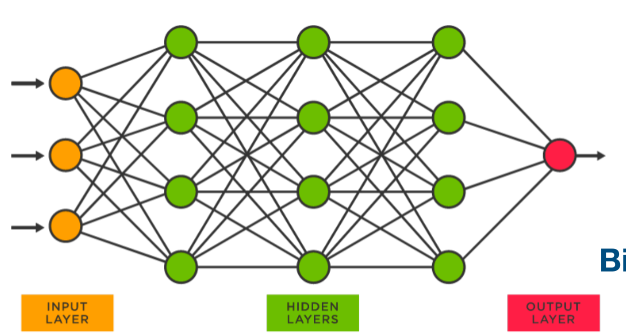
sum over processes
$c = S, B,$ **etc.**

$$\boxed{\frac{p(x_i \,|\, \mu, \alpha)}{p_{ref}(x_i)}} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i \,|\, \alpha) \cdot \nu_c \cdot \boxed{\frac{p_c(x_i)}{p_{ref}(x_i)}} \right]$$

$x \sim$ multi-dimensional kinematic inputs

$x \sim p_c$
$S = 1$

$x \sim p_{ref}$
$S = 0$



$$\underset{\omega}{\text{argmin}} \; \text{L} \longrightarrow \hat{s}(x) = \frac{p_c}{p_{ref} + p_c}(x) \longrightarrow \boxed{\frac{p_c}{p_{ref}}(x) = \frac{\hat{s}(x)}{1.0 - \hat{s}(x)}}$$

Binary Cross-Entropy loss

Two hypothesis: p_c and p_{ref}

Classification NN

"Likelihood ratio trick" or CARL approach [1506.02169]

Many examples in ATLAS - HH4b background estimation, Omnifold, etc.

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S, B,$ etc.

parameter-independent ratio

$$\frac{p(x_i | \mu, \alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha) \cdot f_c(\mu) \cdot \nu_c} \sum_c \left[f_c(\mu) \cdot g_c(x_i | \alpha) \cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **nuisance parameter** α-dependence:

$$g_c(x | \alpha) = \frac{p_c(x | \alpha)}{p_c(x)}$$

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S, B,$ **etc.**

parameter-independent ratio

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **nuisance parameter** α-dependence:

$$g_c(x|\alpha) = \frac{p_c(x|\alpha)}{p_c(x)}$$

Challenging due to the high-dimensionality of $\alpha = (\alpha_m)$

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S, B$, etc.

parameter-independent ratio

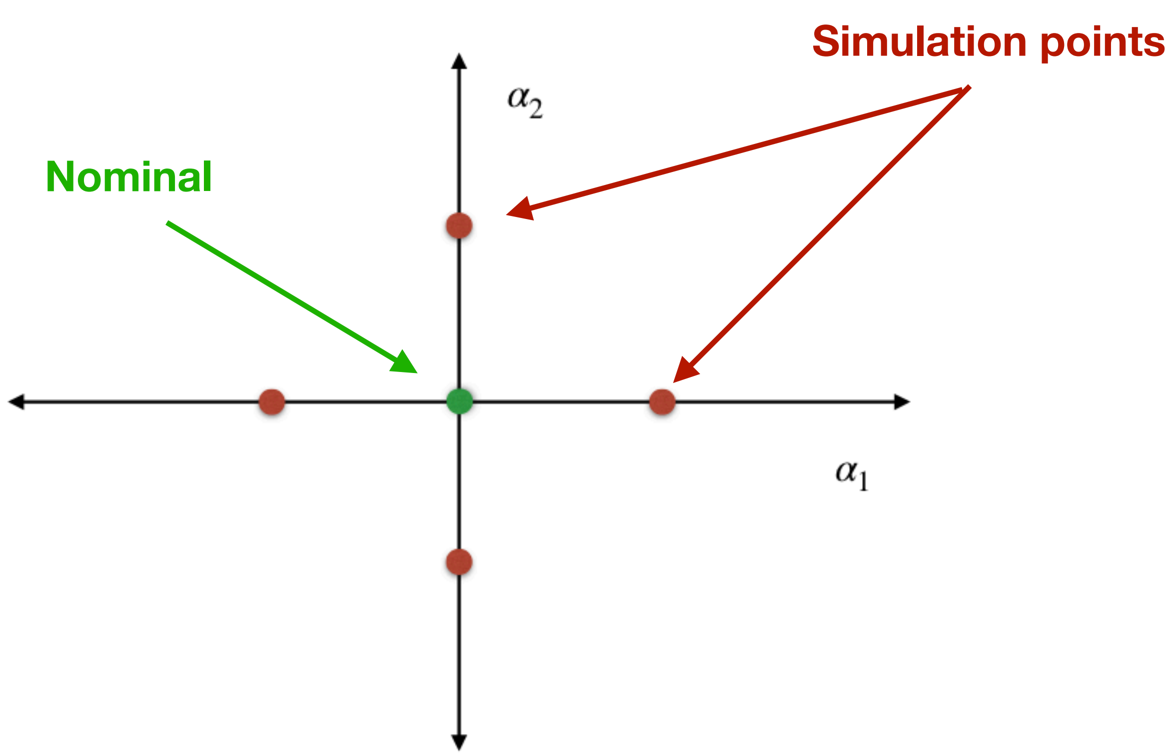
$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c} \sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **nuisance parameter** α-dependence:

$$g_c(x|\alpha) = \frac{p_c(x|\alpha)}{p_c(x)}$$

Challenging due to the high-dimensionality of $\alpha = (\alpha_m)$

Assumption 1:



The effects from the various NPs α_m are orthogonal to each other

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S,\ B,$ **etc.**

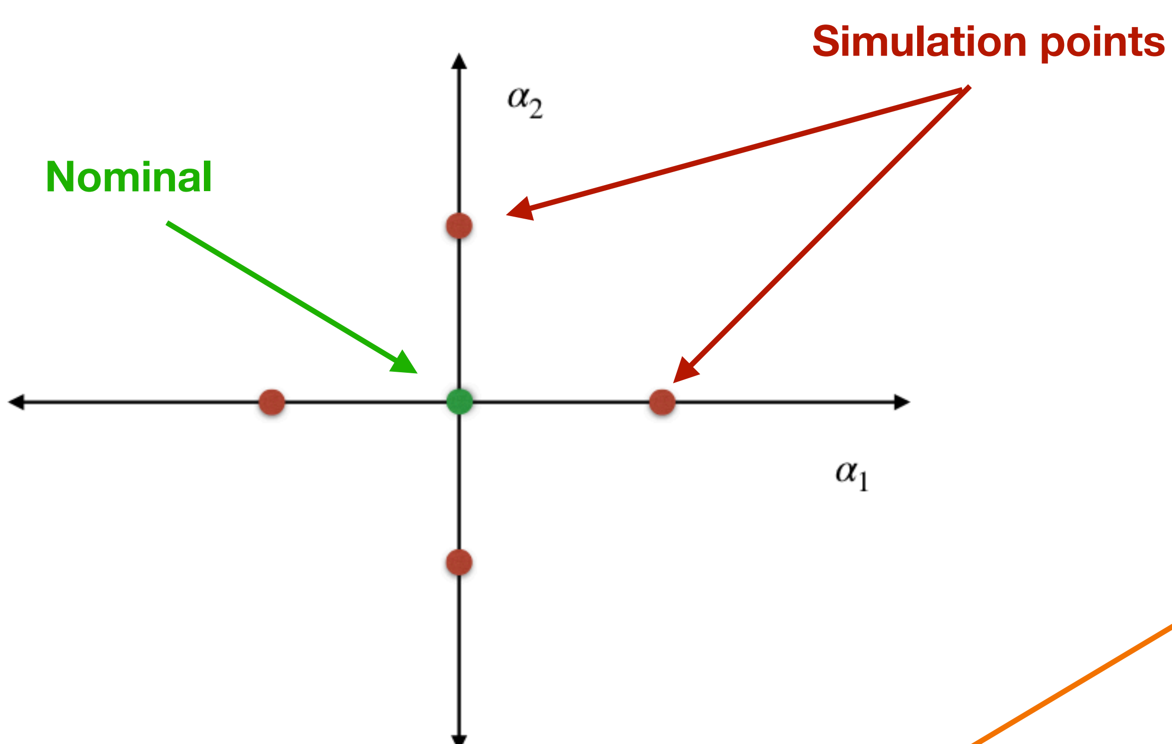
parameter-independent ratio

$$\boxed{\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)}} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c \left[f_c(\mu)\cdot g_c(x_i\,|\,\alpha)\cdot \nu_c \cdot \boxed{\frac{p_c(x_i)}{p_{ref}(x_i)}} \right]$$

Factorized **nuisance parameter** α-dependence:

$$g_c(x\,|\,\alpha) = \frac{p_c(x\,|\,\alpha)}{p_c(x)} = \prod_m \frac{p_c(x\,|\,\alpha_m)}{p_c(x)}$$

Assumption 1:



Often a fair assumption for the systematics model at the LHC

The effects from the various NPs α_m are orthogonal to each other

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S,\ B,$ **etc.**

parameter-independent ratio

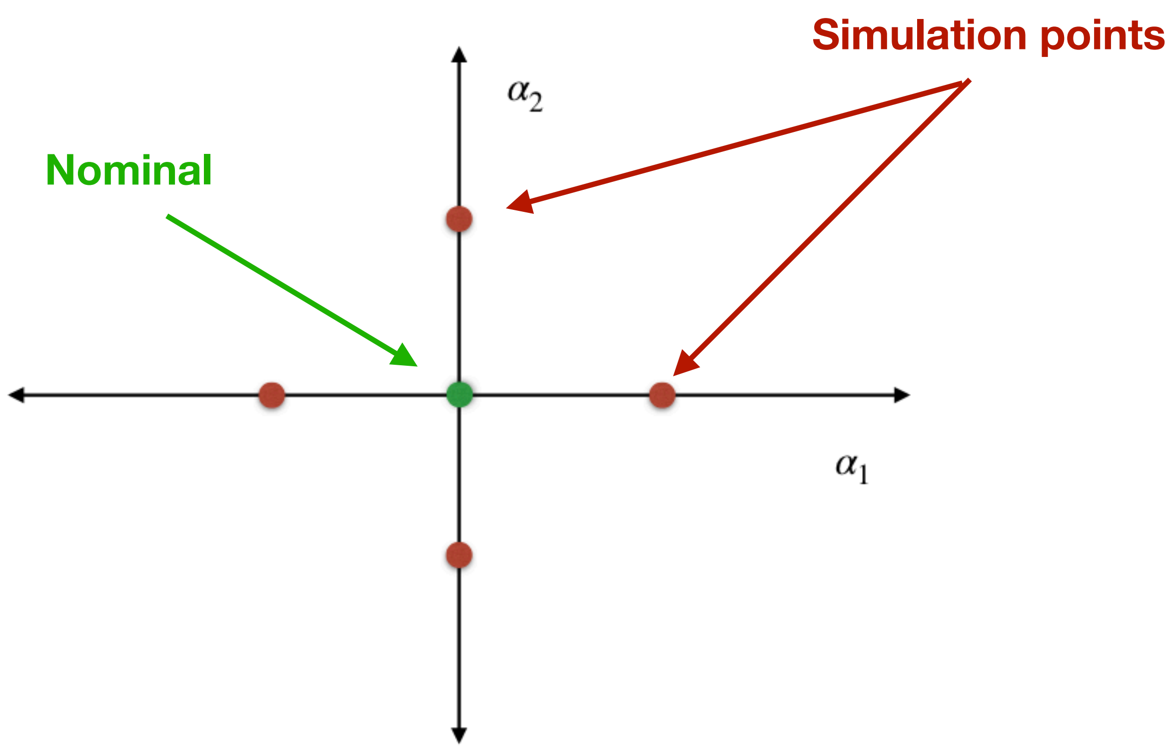
$$\frac{p(x_i\,|\,\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c\left[f_c(\mu)\cdot g_c(x_i\,|\,\alpha)\cdot \nu_c\cdot\frac{p_c(x_i)}{p_{ref}(x_i)}\right]$$

Factorized **nuisance parameter** α-dependence:

$$g_c(x\,|\,\alpha) = \frac{p_c(x\,|\,\alpha)}{p_c(x)} = \prod_m \frac{p_c(x\,|\,\alpha_m)}{p_c(x)}$$

But we also need to estimate these parameterized density ratios

Assumption 1:



The effects from the various NPs α_m are orthogonal to each other

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S, B$, etc.

parameter-independent ratio

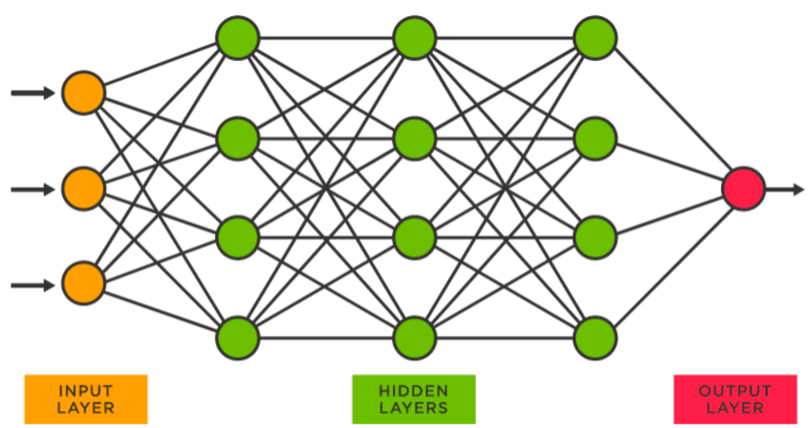
$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c\left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot\nu_c\cdot\frac{p_c(x_i)}{p_{ref}(x_i)}\right]$$

Train parameterized NNs for each α_m

α_m

$x \sim p_c(\alpha_m)$
$S = 1$

$x \sim p_c$
$S = 0$



$$\underset{\omega}{\text{argmin}}\ \text{L} \longrightarrow \frac{p_c(x|\alpha_m)}{p_c(x)}$$

Binary Cross-Entropy loss

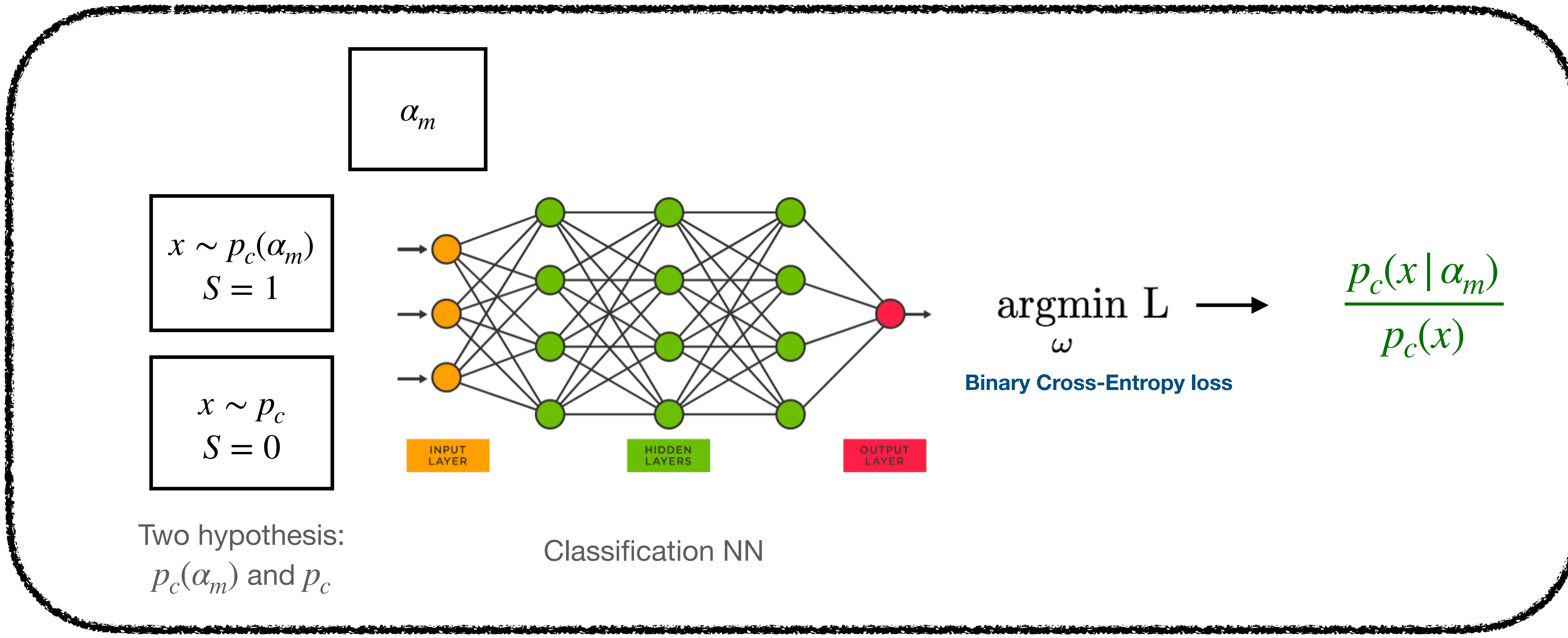
Two hypothesis:
$p_c(\alpha_m)$ and p_c

Classification NN

"CARL" approach

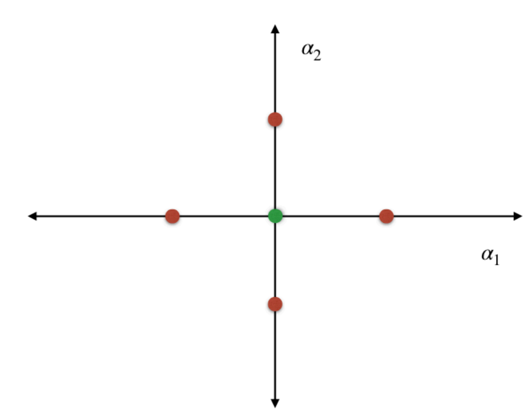
Challenges: Systematic Uncertainties

Train parameterized NNs for each α_m



Challenges:

- **Simulations only available at 3 parameter points -** $\alpha_m^0, \alpha_m^{\pm 1\sigma_a}$
- **Difficult to validate the NN interpolation into phase space with no simulations for testing.**



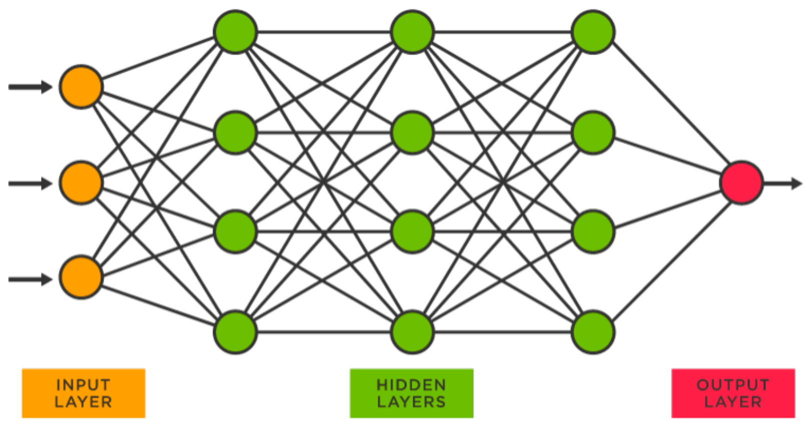
Challenges: Systematic Uncertainties

Solution: train single unparameterized NNs for each $\alpha_m^{\pm 1\sigma_a}$

$x \sim p_c(\alpha_m^{\pm 1\sigma_a})$
$S = 1$

$x \sim p_c$
$S = 0$

Two hypothesis:
$p_c(\alpha_m)$ and p_c



Classification NN

$\underset{\omega}{\text{argmin}}\ \text{L} \longrightarrow \frac{p_c(x \mid \alpha_m^{\pm 1\sigma_a})}{p_c(x)}$

Binary Cross-Entropy loss

Assumption 2:

Semi-analytic approximation

$$\frac{p_c(x|\alpha_m)}{p_c(x)} = \begin{cases} \left(\frac{p_c(x|\alpha_m^{+1\sigma_a})}{p_c(x)}\right)^{\alpha_m} & \alpha_m > 1 \\ 1 + \sum_{i=1}^{6} a_i \boldsymbol{\alpha}_m^i & -1 \leq \alpha_m \leq 1 \\ \left(\frac{p_c(x|\alpha_m^{-1\sigma_a})}{p_c(x)}\right)^{-\alpha_m} & \alpha_m < -1 \end{cases}$$

The α-dependent negative log-likelihood ratio is a smooth parabolic function

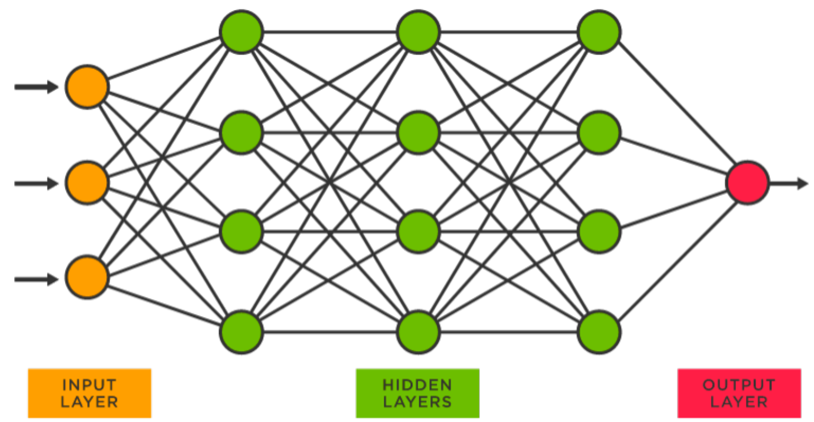
Challenges: Systematic Uncertainties

Solution: train single unparameterized NNs for each $\alpha_m^{\pm 1\sigma_a}$

$x \sim p_c(\alpha_m^{\pm 1\sigma_a})$
$S = 1$

$x \sim p_c$
$S = 0$

Two hypothesis:
$p_c(\alpha_m)$ and p_c



Classification NN

$$\underset{\omega}{\operatorname{argmin}}\ \mathrm{L} \longrightarrow \frac{p_c(x\,|\,\alpha_m^{\pm 1\sigma_a})}{p_c(x)}$$

Binary Cross-Entropy loss

Assumption 2:

Semi-analytic approximation

$$\frac{p_c(x|\alpha_m)}{p_c(x)} = \begin{cases} \left(\dfrac{p_c(x|\alpha_m^{+1\sigma_a})}{p_c(x)}\right)^{\alpha_m} & \alpha_m > 1 \\ 1 + \sum_{i=1}^{6} a_i \boldsymbol{\alpha}_m^i & -1 \le \alpha_m \le 1 \\ \left(\dfrac{p_c(x|\alpha_m^{-1\sigma_a})}{p_c(x)}\right)^{-\alpha_m} & \alpha_m < -1 \end{cases}$$

Combine with → **Assumption 1**

$$g_c(x\,|\,\alpha) = \prod_m \frac{p_c(x\,|\,\alpha_m)}{p_c(x\,|\,\alpha_m^0 = 0)}$$

The α-dependent negative log-likelihood ratio is a smooth parabolic function

Same ideas as proposed HistFactory, extrapolated to NSBI (see backup slides)

Challenges: Systematic Uncertainties

Parameterized per-event ratios

sum over processes $c = S, B,$ etc.

parameter-independent ratio

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Repeat for each systematic, for each process

Assumption 2:

Semi-analytic approximation

$$\frac{p_c(x|\alpha_m)}{p_c(x)} = \begin{cases} \left(\frac{p_c(x|\alpha_m^{+1\sigma_a})}{p_c(x)}\right)^{\alpha_m} & \alpha_m > 1 \\ 1 + \sum_{i=1}^{6} a_i \boldsymbol{\alpha}_m^i & -1 \leq \alpha_m \leq 1 \\ \left(\frac{p_c(x|\alpha_m^{-1\sigma_a})}{p_c(x)}\right)^{-\alpha_m} & \alpha_m < -1 \end{cases}$$

Combine with Assumption 1 →

$$g_c(x|\alpha) = \prod_m \frac{p_c(x|\alpha_m)}{p_c(x|\alpha_m^0 = 0)}$$

The α-dependent negative log-likelihood ratio is a smooth parabolic function

Same ideas as proposed HistFactory, extrapolated to NSBI (see backup slides)

Results using NSBI

The NSBI approach learns everything, including the **parameter scaling and thus the full interference effects**

$$\frac{p(x\,|\,\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_{ref}(x)}} + \sqrt{\mu} \cdot \nu_I \cdot \boxed{\frac{p_I(x)}{p_{ref}(x)}} + \nu_B \cdot \boxed{\frac{p_B(x)}{p_{ref}(x)}} + \nu_{NI} \cdot \boxed{\frac{p_{NI}(x)}{p_{ref}(x)}}$$

Four parameter-independent ratios are trained

(suppressing the α-terms for brevity)

Results using NSBI

The NSBI approach learns everything, including the **parameter scaling and thus the full interference effects**

$$\frac{p(x\,|\,\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_{ref}(x)}} + \sqrt{\mu} \cdot \nu_I \cdot \boxed{\frac{p_I(x)}{p_{ref}(x)}} + \nu_B \cdot \boxed{\frac{p_B(x)}{p_{ref}(x)}} + \nu_{NI} \cdot \boxed{\frac{p_{NI}(x)}{p_{ref}(x)}}$$

Four parameter-independent ratios are trained

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i\,|\,\mu)}{p(x_i\,|\,\hat{\mu})}$$

No "fixed" S/B discriminant - asymptotic optimality throughout μ space.

Additional sensitivity from unbinned nature

Results using NSBI

The NSBI approach learns everything, including the **parameter scaling and thus the full interference effects**

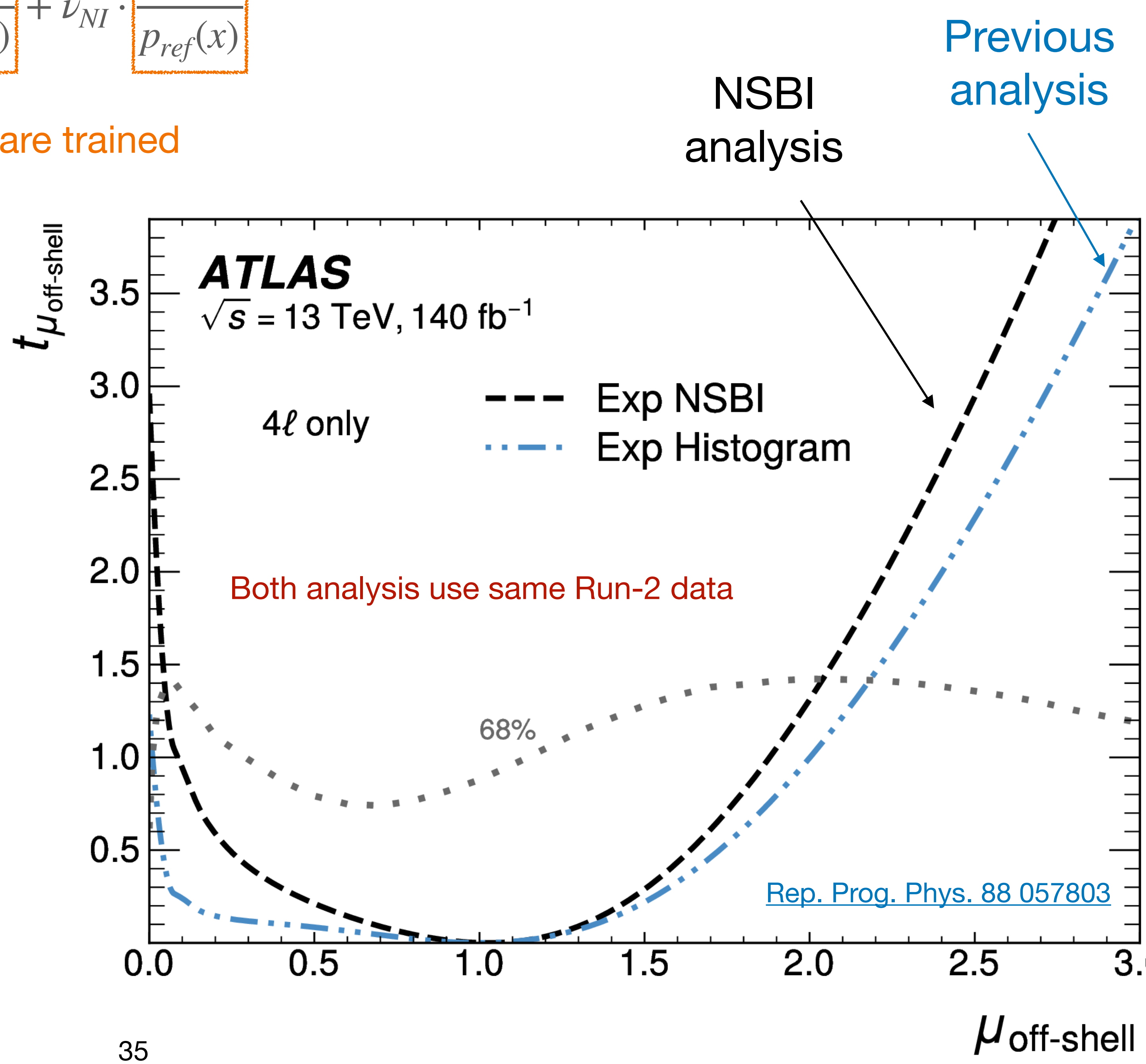
$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_{ref}(x)}} + \sqrt{\mu} \cdot \nu_I \cdot \boxed{\frac{p_I(x)}{p_{ref}(x)}} + \nu_B \cdot \boxed{\frac{p_B(x)}{p_{ref}(x)}} + \nu_{NI} \cdot \boxed{\frac{p_{NI}(x)}{p_{ref}(x)}}$$

Four parameter-independent ratios are trained

$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i|\mu)}{p(x_i|\hat{\mu})}$$

No "fixed" S/B discriminant - asymptotic optimality throughout μ space.

Additional sensitivity from unbinned nature



Results using NSBI

The NSBI approach learns everything, including the **parameter scaling and thus the full interference effects**

$$\frac{p(x|\mu)}{p_B(x)} \sim \mu \cdot \nu_S \cdot \boxed{\frac{p_S(x)}{p_{ref}(x)}} + \sqrt{\mu} \cdot \nu_I \cdot \boxed{\frac{p_I(x)}{p_{ref}(x)}} + \nu_B \cdot \boxed{\frac{p_B(x)}{p_{ref}(x)}} + \nu_{NI} \cdot \boxed{\frac{p_{NI}(x)}{p_{ref}(x)}}$$

Four parameter-independent ratios are trained

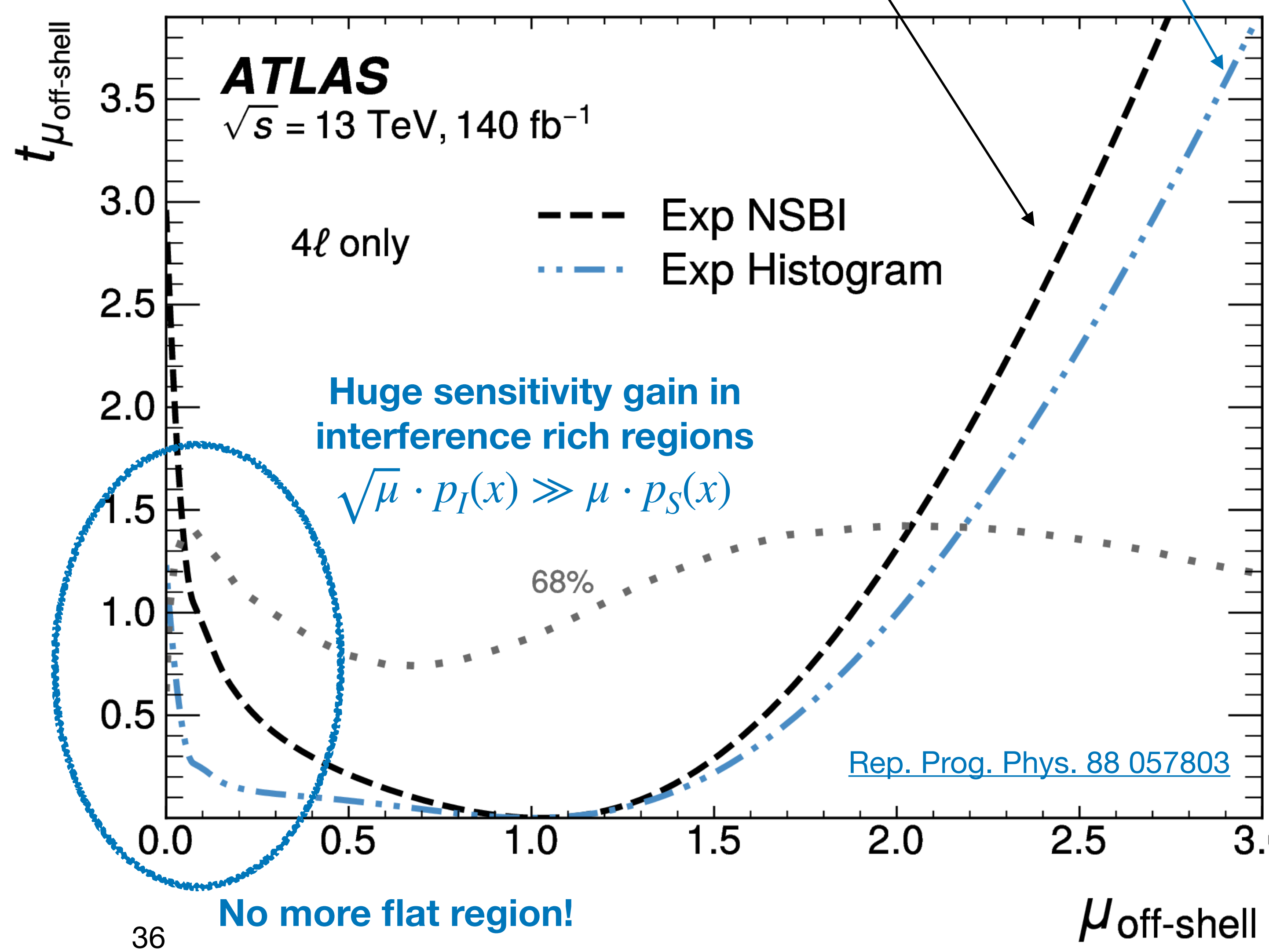
$$t_\mu \sim -2 \cdot \sum_{i=1}^{N_{obs}} \log \frac{p(x_i|\mu)}{p(x_i|\hat{\mu})}$$

No "fixed" S/B discriminant - asymptotic optimality throughout μ space.

Additional sensitivity from unbinned nature

NSBI analysis

Previous analysis



Huge sensitivity gain in interference rich regions

$\sqrt{\mu} \cdot p_I(x) \gg \mu \cdot p_S(x)$

No more flat region!

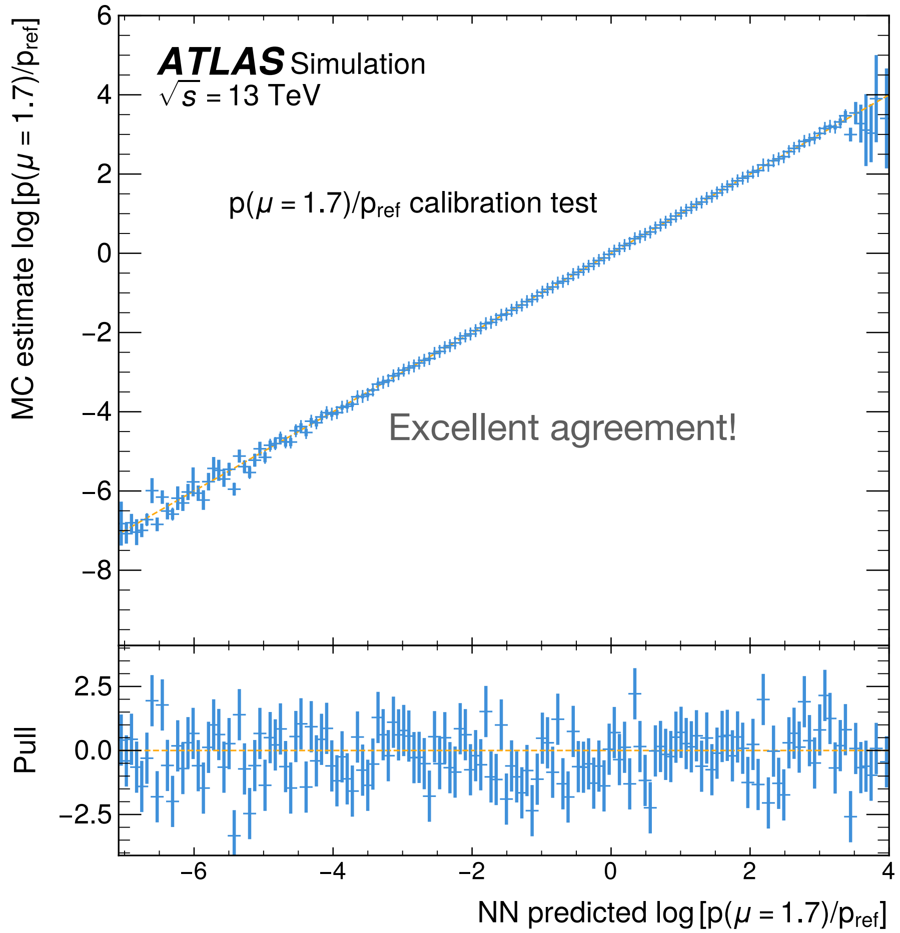
Monte Carlo Diagnostics

The NN ratios are meticulously trained to be **true representations of the density ratios**

Does the NN output correspond to real probabilities?

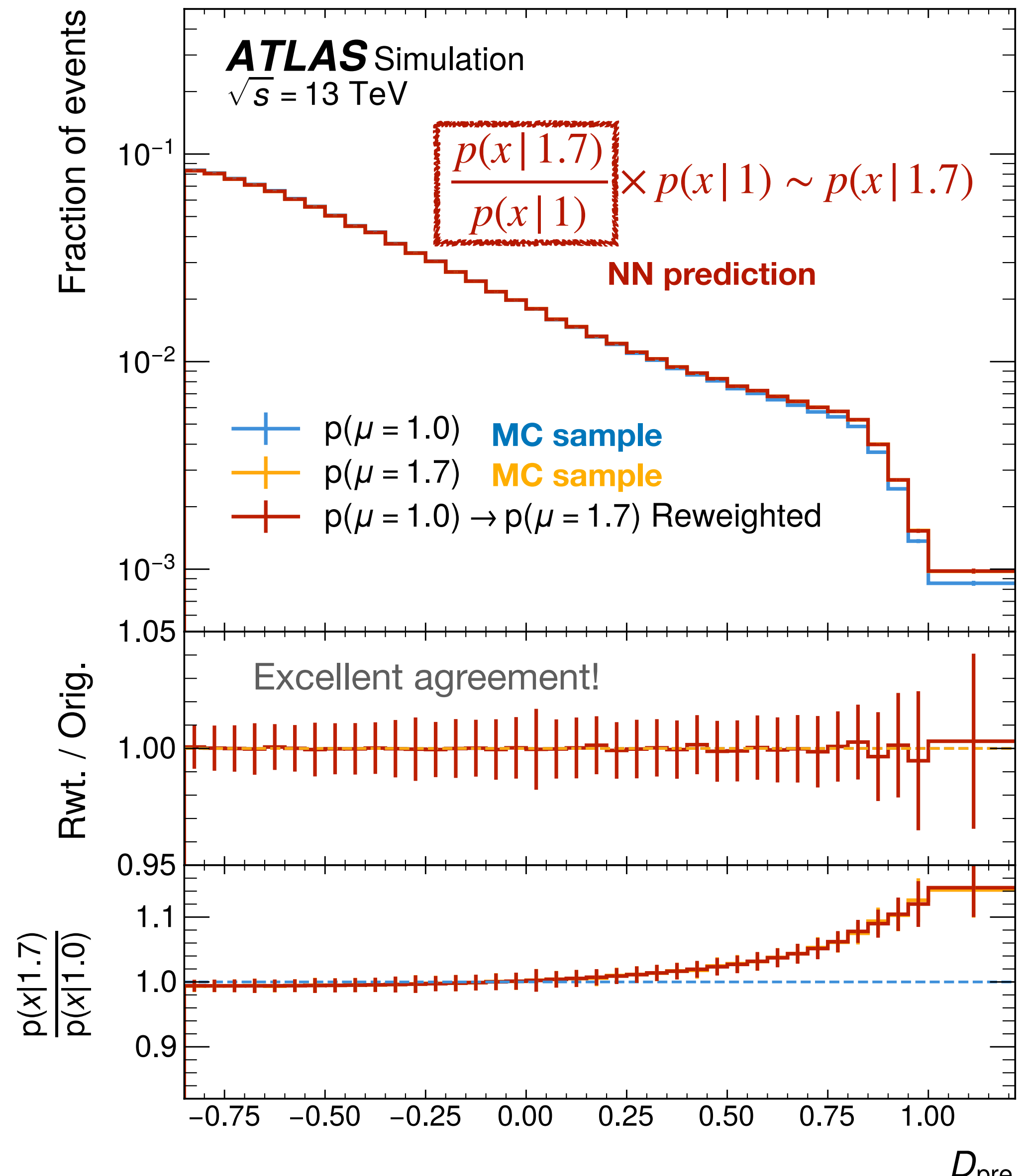
$$\log \frac{p(x|\mu)}{p_{ref}(x)}(x) \leftrightarrow \log \frac{N^{bin}(x|\mu)}{N^{bin}_{ref}(x)} \ ?$$

MC estimate $\log \frac{N^{bin}(x|\mu)}{N^{bin}_{ref}(x)}$



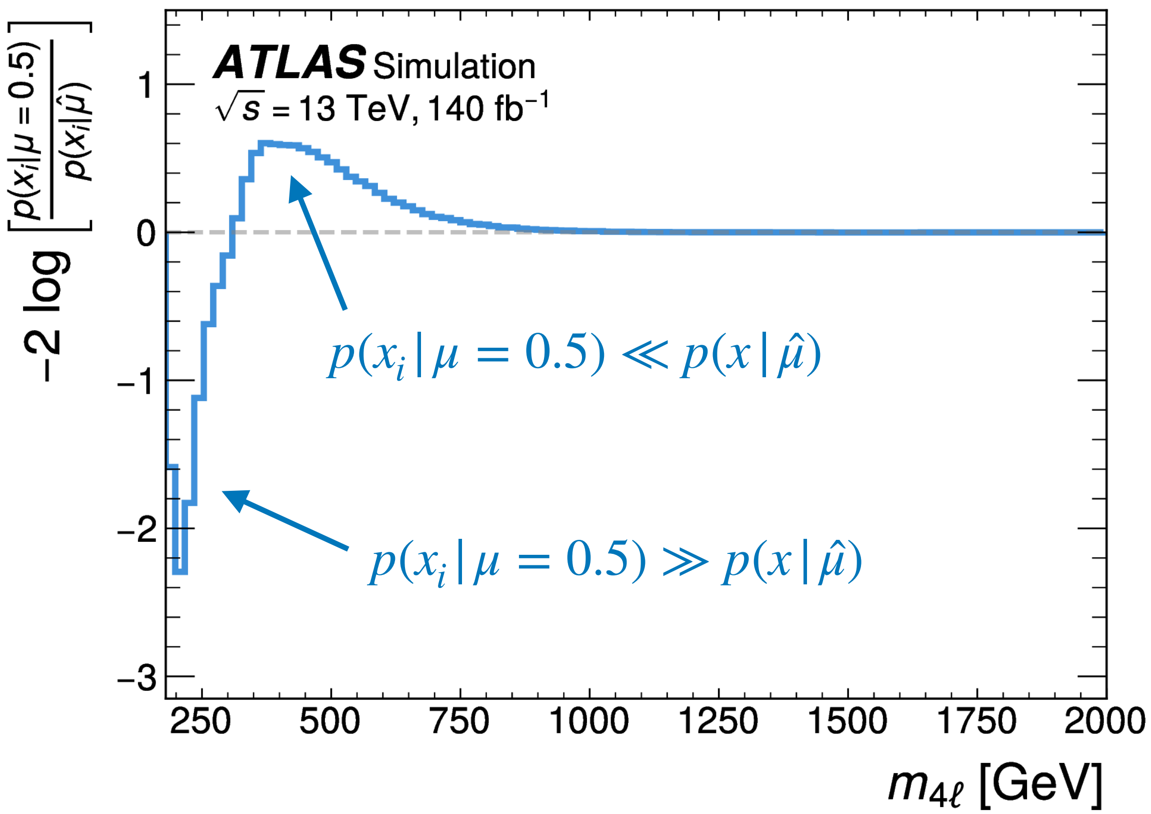
NN predicted $\log \frac{p(x|\mu)}{p_{ref}(x)}(x)$

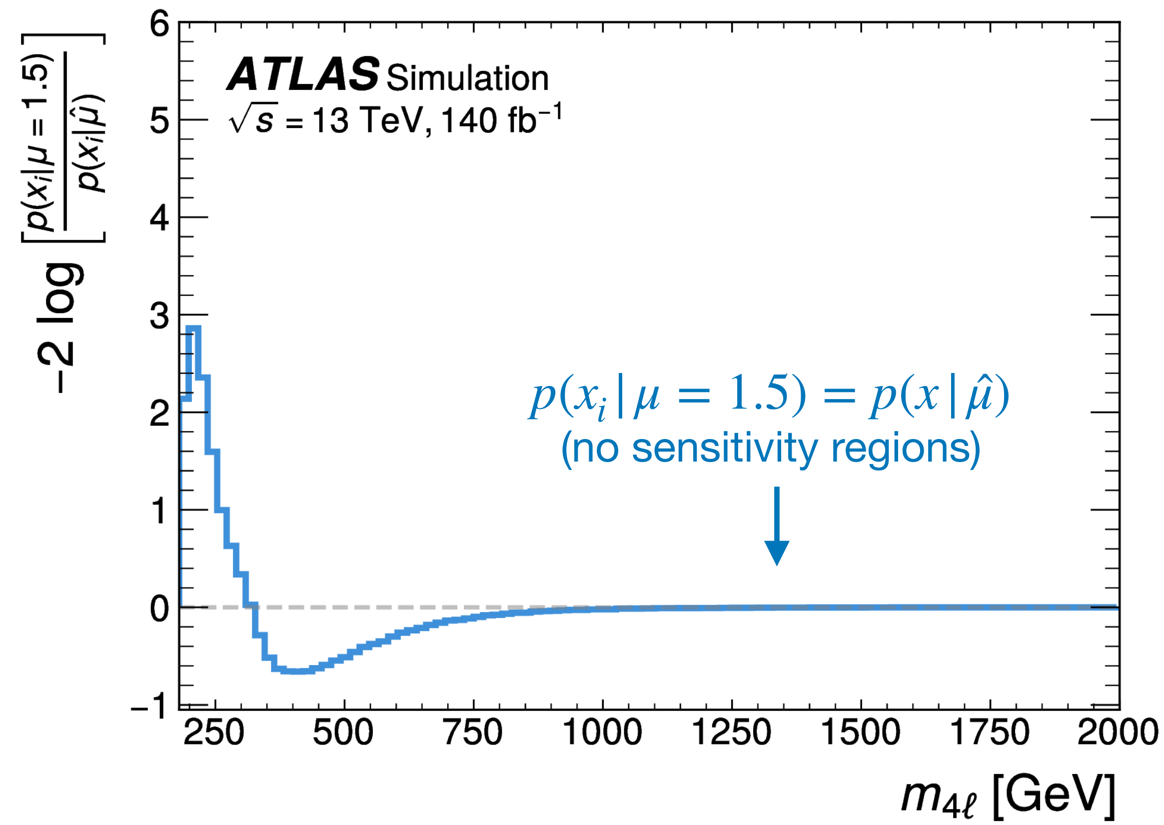
Do the ratios capture the full un-biased dependence of the multi-dimensional feature space x ?



High-level multivariate observable

Where does the sensitivity come from?



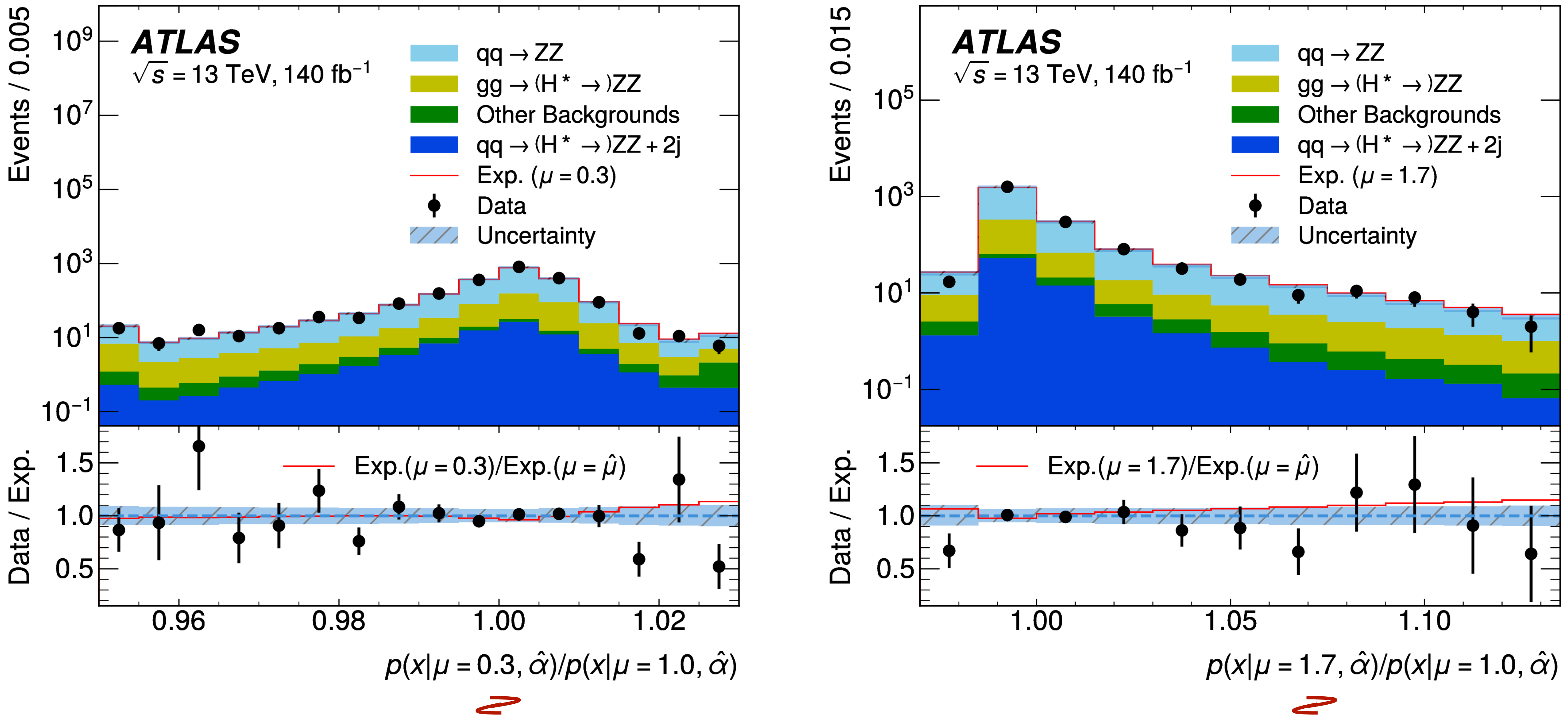


The per-event negative log-likelihood ratio allows a probe to identify phase space regions that contribute to the final analysis senstivity.

This allows to identify phase space regions that need robust modeling from Monte Carlo samples.

Real Data Diagnostics

A rigorous data-MC comparison is performed using the parameterized density ratios



Check agreement as a function of any parameter μ value

Building Frequentist Confidence Intervals

- In settings like the off-shell Higgs boson that do not follow Wald approximation:

$$-2\ln\lambda(\mu) = \frac{(\mu-\hat{\mu})^2}{\sigma^2}$$ ✘

arXiv: 1007.1727

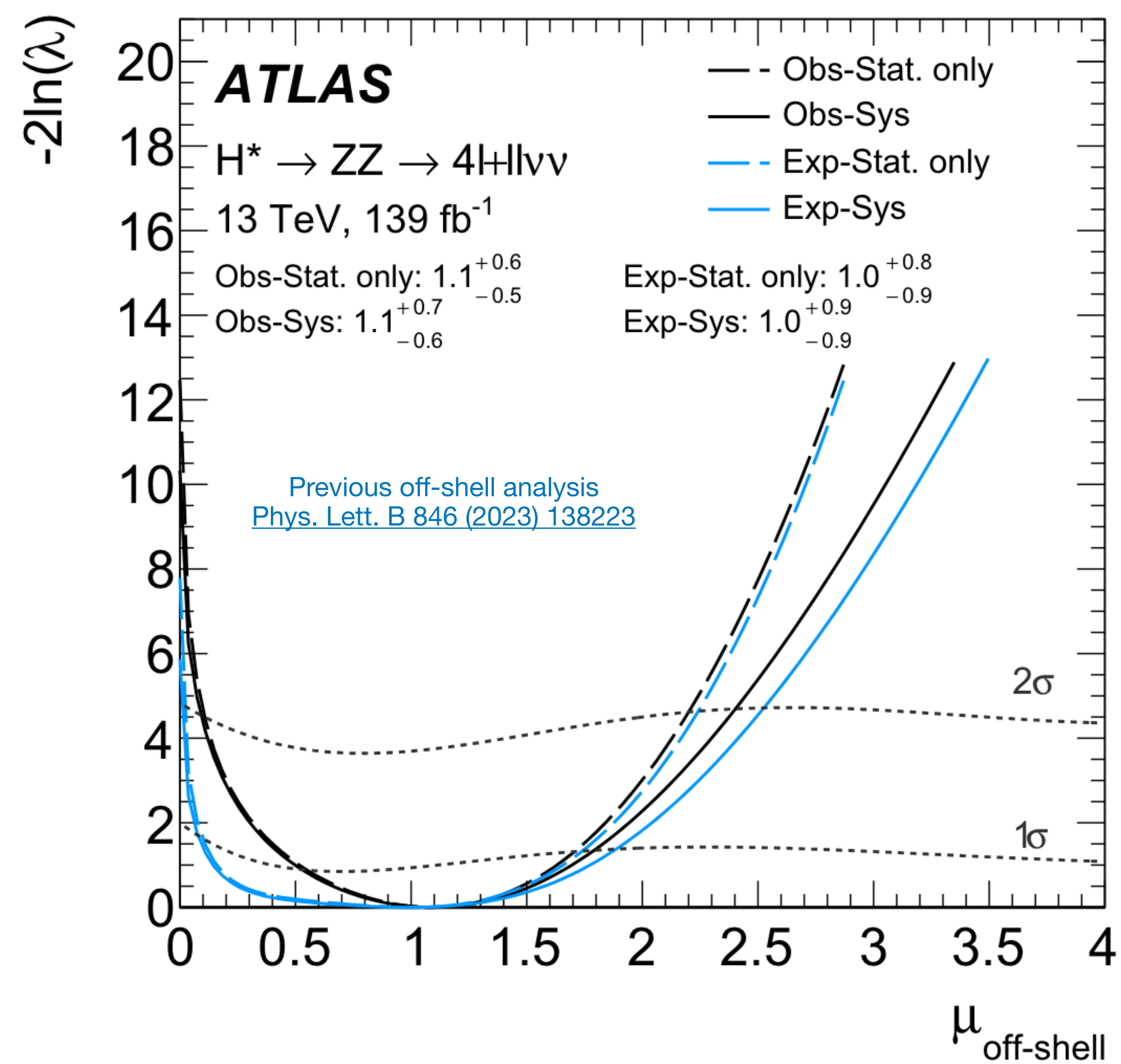
- **Neyman construction is essential.**

- But standard LHC techniques like Poisson PDF sampling cannot work directly.

- This is because the NSBI technique presented here does not have a PDF $p(x\,|\,\mu,\alpha)$ to sample pseudo-data from - only the density ratios:

$$\frac{p(x\,|\,\mu,\alpha)}{p_{ref}(x)}$$

Previous Histogram-based $H^* \to ZZ$ measurement



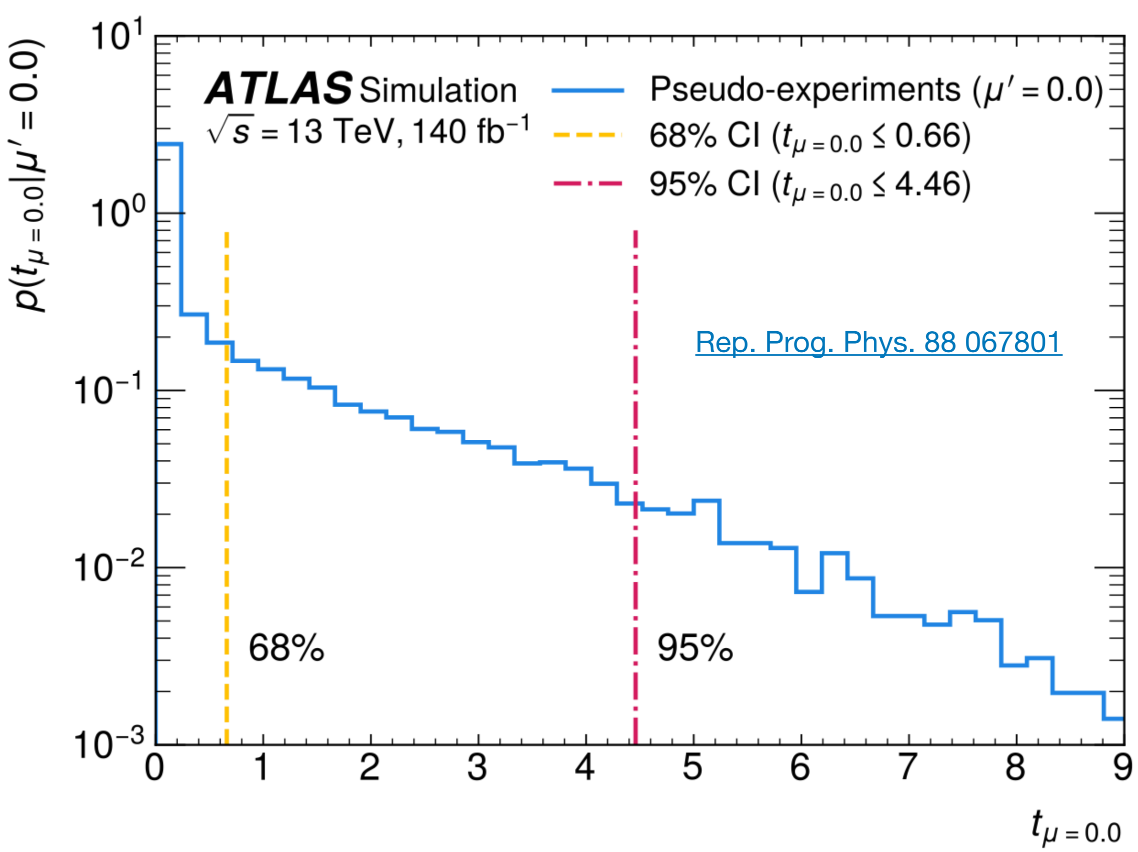
Neyman Construction for NSBI

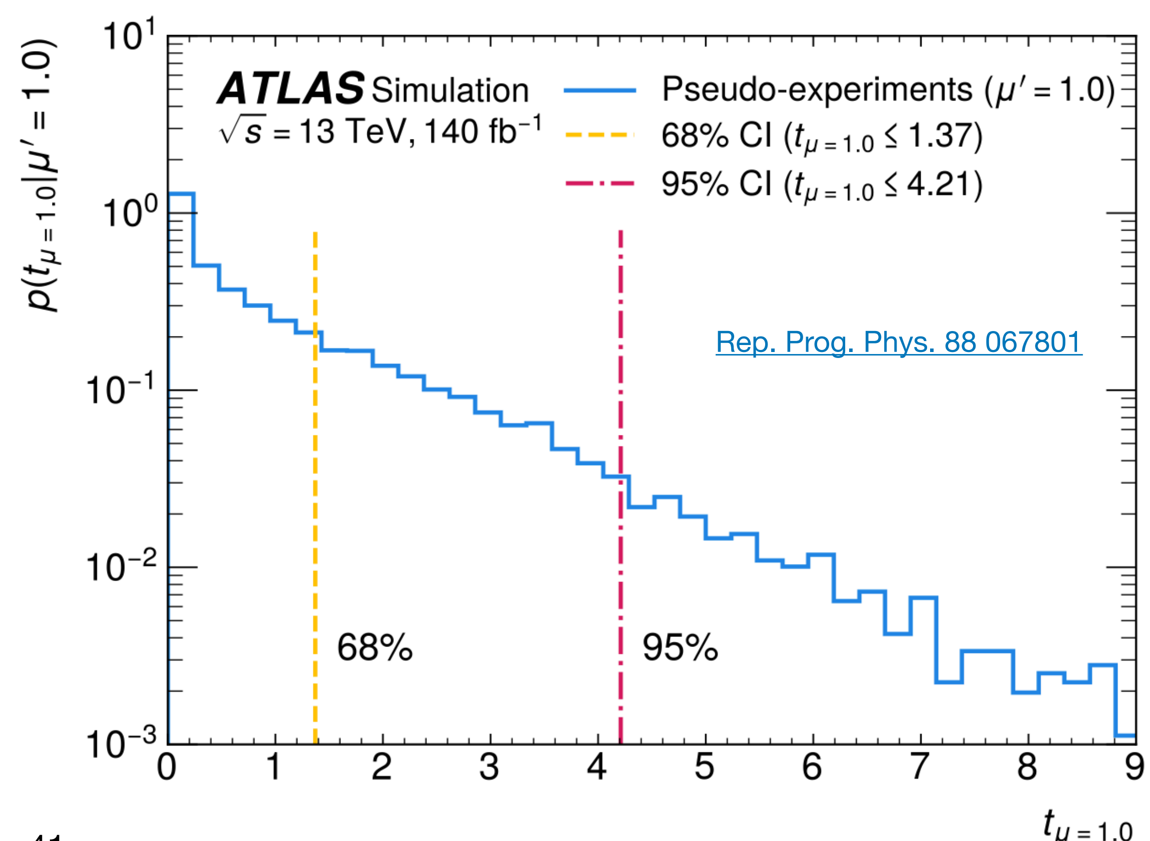
- The trained density ratios are used to create unbiased Asimov samples with MC weights w_A for any value of the μ, α parameter space:

$$w_A(x\,|\,\mu_{truth}, \alpha) = \frac{\nu(\mu, \alpha)}{\nu_{ref}} \cdot \frac{p(x\,|\,\mu_{truth}, \alpha)}{p_{ref}(x)} \cdot w_{ref}(x)$$

MC weights of reference sample

- Pseudo-experiments are then sampled using the Poisson bootstrap method - $w_{pseudo-data}(x) = \text{Poisson}(\ w_A(x\,|\,\mu_{truth}, \alpha)\ )$.

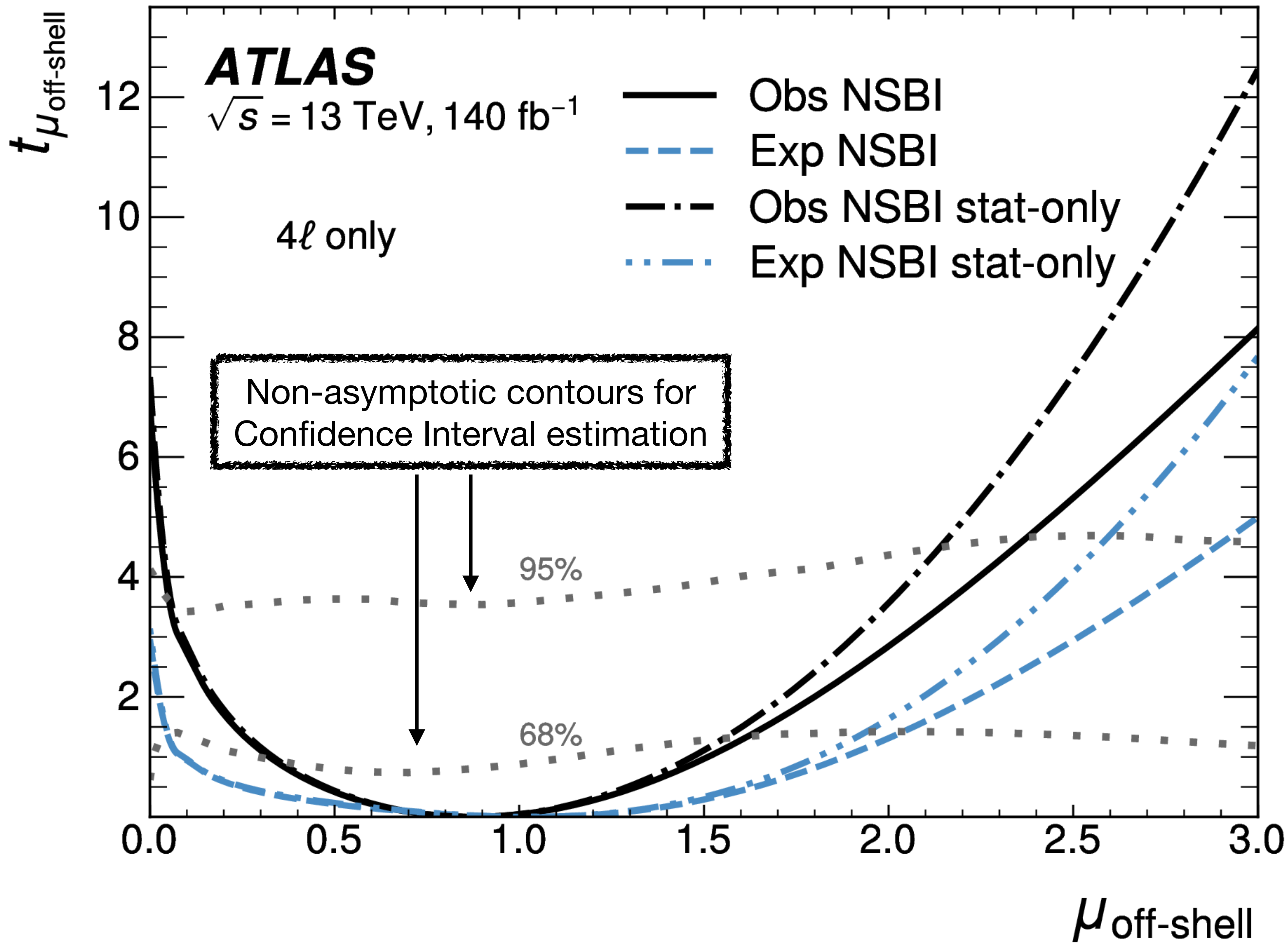


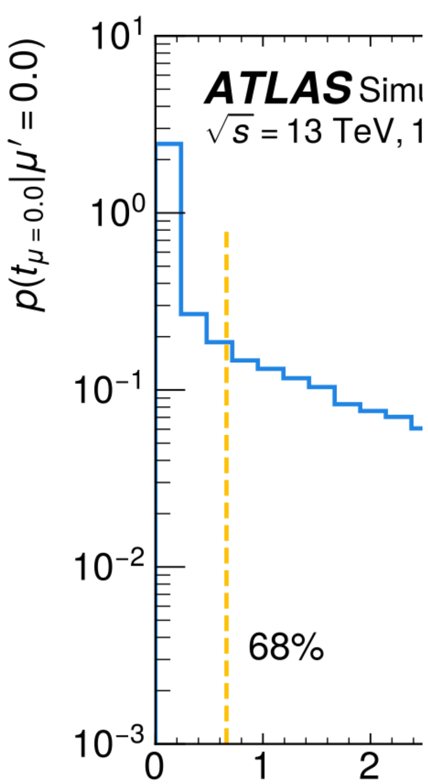


Neyman Construction for NSBI

- The trained density ratios are used to create unbiased Asimov samples with MC weights w_A for any value of the μ, α parameter space:

- Pseudo-experim
$w_{pseudo-data}(x)$:





$t_{\mu=0.0}$

$t_{\mu=1.0}$

Conclusion and Outlook

- An implementation of NSBI was presented focused on building likelihood ratios as a function of complex, large-dimensional parameter spaces using well-motivated approximations.

- **The NSBI approach presented in this talk has broad applicability across LHC analysis** - particularly effective when the likelihood model is non-linear in the parameter of interest and when multi-dimensional information is needed for extra precision.

- The various conceptual and computational developments have been done and published in these companion papers by ATLAS:

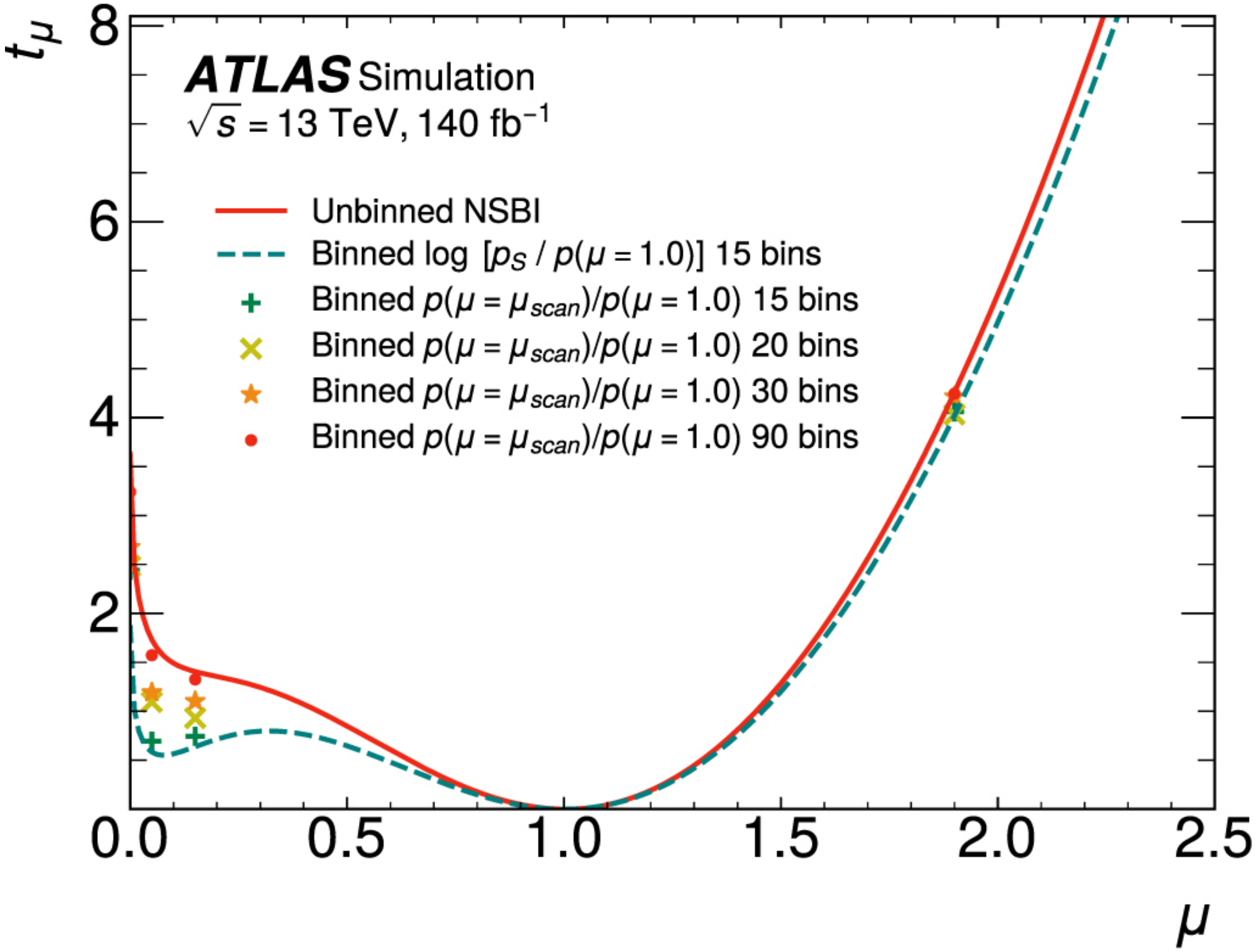
  - Rep. Prog. Phys. 88 067801 [General NSBI method presented in this talk]

  - Rep. Prog. Phys. 88 057803 [Application to off-shell Higgs boson and Higgs boson width measurement with the ATLAS experiment]

Backup

Parameterized Observables and Unbinning

Rep. Prog. Phys. 88 067801



The improved sensitivity from using the NSBI approach is a result of:

- using **parameterized information** for the hypothesis testing
- and doing an **unbinned fit**

Uncertainty Parameterization

$$\frac{p(x_i|\mu,\alpha)}{p_{ref}(x_i)} = \frac{1}{\sum_c G_c(\alpha)\cdot f_c(\mu)\cdot \nu_c}\sum_c \left[f_c(\mu)\cdot g_c(x_i|\alpha)\cdot \nu_c \cdot \frac{p_c(x_i)}{p_{ref}(x_i)} \right]$$

Factorized **yield** α-dependence:

$$G_c(\alpha) = \prod_k \frac{\nu_c(\alpha_k)}{\nu_c}$$

with $\nu_c(\alpha_k)/\nu_c$ estimated using **analytic interpolation techniques:**

Available from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

$$\frac{\nu_c(\alpha_k)}{\nu_c} = \begin{cases} \left(\boxed{\frac{\nu_c(\alpha_k^+)}{\nu_c}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1, \\ \left(\boxed{\frac{\nu_c(\alpha_k^-)}{\nu_c}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Per-event analog of standard techniques →

Factorized **per-event** α-dependence:

$$g_c(x|\alpha) = \prod_k \frac{p_c(x|\alpha_k)}{p_c(x)}$$

with $p_c(x|\alpha_k)/p_c(x)$ estimated using a **mix of NNs and analytic interpolation techniques:**

Density ratios trained using NNs from simulations at $\alpha_k = 0,\ \alpha_k^+,\ \alpha_k^-$

$$\frac{p_c(x|\alpha_k)}{p_c(x)} = \begin{cases} \left(\boxed{\frac{p_c(x|\alpha_k^+)}{p_c(x)}}\right)^{\alpha_k} & \alpha_k > 1 \\ 1 + \sum_{n=1}^{6} c_n \alpha_k^n & -1 \le \alpha_k \le 1 \cdot \\ \left(\boxed{\frac{p_c(x|\alpha_k^-)}{p_c(x)}}\right)^{-\alpha_k} & \alpha_k < -1 \end{cases}$$

Ref: HistFactory

Full workflow of the NSBI Analysis